Freedom and Experience

Self-Determination without Illusions

Kevin Magill

FREEDOM AND EXPERIENCE

Freedom and Experience

Self-Determination without Illusions

Kevin Magill
Lecturer in Philosophy
University of Wolverhampton

This edition published by the author as Open Access, 2016.

Originally ublished in Great Britain 1997 by MACMILLAN PRESS LTD, Houndmills, Basingstoke, Hampshire RG2 I 6XS.

A catalogue record for this book is available from the British Library.

ISBN 0-333-63453-5

Originally published in the United States of America 1997 by
ST. MARTIN'S PRESS, INC.,
Scholarly and Reference Division,
I75 Fifth Avenue, New York, N.Y. I0010

ISBN 0-312-16474--2

Library of Congress Cataloging-in-Publication Data Magill, Kevin, 1959-
Freedom and experience: self-determination without illusions *I* Kevin Magill.
p. cm.
Includes bibliographical references and index.
ISBN 0-312-16474--2 (cloth)
I. Free will and determinism. **I.** Title.
BJI460.M34 1996 96-30255
I23'.5---dc20 CIP

10 9 8 7 6 5 4 3 2 I
06 05 04 03 02 CJ 00 99 98 97

Printed and bound in Great Britain by
Antony Rowe Ltd, Chippenham, Wiltshire

To Mary and Danny, my mother and father, for everything

Contents

Preface

This book is about the nature of free will and agency. It belongs to a long philosophical tradition that has it that free will and determinism are compatible (the *compatibilist* tradition). Opponents of this tradition (*incompatibilists*) have argued that if determinism is true, then our experiences of being free to act and decide in various ways, as well as our beliefs about moral responsibility, desert and punishment, are inherently illusory. Several contemporary philosophers, impressed by the longevity and apparent intractability of the argument about free will and determinism, have suggested that the argument is irresolvable on its own terms, either because there are no settled meanings of *free will, moral responsibility* and related expressions or because our beliefs about free will and agency are incoherent and fuelled by conflicting images and intuitions. If such suggestions are well founded, then our beliefs about free will must be partly or wholly illusory. In contrast to such suggestions, this book argues that the intractability of the traditional argument about free will and determinism has had two principal sources. The first source is the mistaken idea that our practices of punishing and holding people responsible call for a moral and metaphysical justification. The second source has been a lack of attention to the contents and limits of the experiences that shape our understanding of free will and agency. [1]

Chapter 1 argues that the traditional argument about the meanings and implications of *free will, moral responsibility, can* and *could have*, and *necessity* has resulted in stalemate, and considers the claim that this has arisen because we are subject to contradictory and illusory metaphysical attitudes about the initiation of actions. Chapter 2 follows Peter Strawson's 'Freedom and Resentment'[2] in arguing that our practices of holding people responsible, and praising, blaming and punishing them, are part of the general framework of human life and do not stand in need of a general justification, but goes on to point out that this still leaves us with practical difficulties about whether to blame or to try to understand particular instances of wrongdoing. Chapter 3 defends the claim that free will consists in being able to do what you really want because it is what you really want. Chapter 4 examines the argument that determinism would render our beliefs and experiences of making decisions illusory, by considering whether it would be possible to experience ourselves as being caused to make the decisions we do. Chapter 5 looks at what it is that distinguishes actions from involuntary behaviour and whether we are agents such as we

take ourselves to be. Chapter 6 investigates what is required for fully human agency of the kind we take ourselves to have in being able to judge between competing motives for action and in being able to act on such judgements.

The truth of determinism is assumed throughout the book and no at-tempt is made to argue the case in favour of it. One reason for this is that a persuasive case for determinism has already been made elsewhere, to which I have nothing useful to add.[3] A further reason is that in common with many others I do not see how there can be free will without deter-minism. This also has been largely assumed without argument. Again, the case against contemporary indeterministic accounts of free will has been well made elsewhere,[4] but more importantly, what I try to show is that what has led many philosophers to resist the idea that free will is compat-ible with determinism are various experiences of acting and deciding, which compatibilists have largely failed to account for adequately but which are consistent with determinism. If my claims about the sources of incom-patibilist worries are correct, and if I am right in arguing that those worries can be met without recourse to indeterminism, then any refutation of in-determinist accounts of free will will be otiose.

Chapters 2, 4 and 6 are distant descendants of my doctoral thesis, which was submitted to the University of London in 1993. A much earlier ver-sion of Chapter 4 was presented as a paper at the first European Congress of Analytic Philosophy in Aix-en-Provence in 1993. I am grateful to the University of Wolverhampton for a sabbatical semester in 1994, during which the first draft of the book was written. Both the arguments and the structure of the book have benefited greatly from the suggestions and criticisms of my PhD supervisor, Professor Ted Honderich, to whom I am indebted. I am also grateful to several readers who commented on all or part of the manuscript and its ancestors. They are Mark Bernstein, John Bishop, David Cockburn, Meena Dhanda, Richard Double, Danny Gold-stick, Bob Kane, Al Mele, Adam Morton, Galen Strawson, Robin Taylor and David Velleman. I am very grateful to Kimberly Hutchings and Moya Lloyd, who checked the manuscript and helped to eliminate a large number of errors and inaccuracies, to Valery Rose and Jocelyn Stockley for their editorial work, to Mike Cunningham who read the proofs and to Annabelle Buckley for her advice and assistance. Finally, I would like to thank my parents, to whom this book is dedicated, for their unfailing support and encouragement.

KEVIN MAGILL

k.magill@outlook.com
https://wlv.academia.edu/KevinMagill.

1 Are the Problems of Free Will Resolvable?

Most of us are struck from time to time by recollections of past deeds and accomplishments. Some are of achievements and feats of ingenuity, or of times in which we have given help and shown kindness to others. They give cause for satisfaction and pride. At other times we remember acts of meanness, spite and indifference, or of indolence and failure, and are brought to see ourselves in quite a different light. Such memories burden us with shame and regret, and responsibility for hurt done to those we care about. The burden of regret can lead us to turn our attention to the future, in which we may resolve to be kinder and better, and to put past failures behind us.

Reflections about past and future actions, and the feelings they produce, are human, and no matter how troubling and burdensome they can sometimes be to us, it is difficult to see how we could ever come to give them up. And yet, many have questioned whether we are ever justified in our feelings of responsibility, pride or shame about our own actions, or in the congratulations and condemnations we heap on others. To believe ourselves to be responsible for our actions (or to hold others responsible for theirs) is to assume that we can (sometimes) act freely, and thus that we really have a choice about what we do. A more basic assumption than this is that our doings and strivings are of a fundamentally different order (rather than merely differing in degree of mechanical complexity) from the behaviour of inanimate natural phenomena. A familiar philosophical response to these assumptions is that they are rooted in superstition or illusion. What we now know (as our ancestors, it seems, did not) is that we are physical creatures; no more than flesh and blood and capable of no more than what flesh and blood can allow. Like all physical creatures our thoughts and behaviour are subject to natural laws. Our supposed triumphs, just like our supposed failures, are all caused by earlier states and events, themselves caused by yet earlier states and events, and could not have failed to happen as they did: the thesis of determinism. For as long as we continue to see ourselves as responsible for things that could not have happened otherwise, it is argued, we will be enmeshed in illusion.

But ought we to doubt that we are responsible for what we do or that we have a capacity for acting freely? For it surely cannot be denied that human beings do things, and often enough without compulsion. According

to compatibilists, this is, more or less, what we mean when we say that a person has acted freely or that she is responsible for what she has done. The definition might be elaborated so that we are not obliged to say that actions caused by addictions and neuroses are free, and also that there is some kind of principled connection between an agent's free actions and what she wants (or most wants), but in essence, it is argued, the definition is sound. And, as many compatibilists have pointed out, a person may act without compulsion even though her actions are governed by causal laws and could, with sufficient knowledge about those laws and the causal antecedents of the actions, be accurately predicted. Incompatibilists have insisted that if our actions are subject to causal laws, they must be physically necessitated, and this entails that we never really have any choice about them and cannot fairly be held responsible for them. Some have also argued that what we mean by ***actions*** are events that are caused by us as agents, and if we say instead that actions are caused by antecedent events, even events within us, we can no longer persist in thinking of ourselves as having any role as agents in the production of our actions.

Is there any way of resolving this disagreement, so that we can say, once and for all, whether our assumptions about agency, free will and responsibility are justified or illusory? The argument between compatibilism and incompatibilism is an old one and seems no closer to a resolution now than at any time in the past. The unyielding character of the traditional argument and the convictions that inform it can be seen from the following four quotations:

> For what is meant by liberty, when applied to voluntary actions? We cannot surely mean that actions have so little connexion with motives, inclinations, and circumstances, that one does not follow with a certain degree of uniformity from the other, and that one affords no inference by which we can conclude the existence of the other. For these are plain and acknowledged matters of fact. By liberty, then, we can only mean ***a power of acting or not acting, according to the determinations of the will;*** that is, if we choose to remain at rest, we may; if we choose to move, we also may. (David Hume[1])

> Suppose I say of a man who has committed a theft that this act, by the natural law of causality, is a necessary result of the determining ground existing in the preceding time and that it was therefore impossible that it could have not been done. How, then, can judgment according to the moral law make any change in it? And how can it be supposed that it still could have been left undone because the law says that it should

> have been left undone? ... It is a wretched subterfuge to seek an escape in the supposition that the *kind* of determining grounds of his causality according to natural law agrees with a comparative concept of freedom. According to this concept, what is sometimes called 'free effect' is that of which the determining natural cause is internal to the acting thing. For example, that which a projectile performs when it is in free motion is called by the name 'freedom' because it is not pushed by anything external while it is in flight ... it would in essence be no better than the freedom of a turnspit, which when once wound up also carries out its motions of itself. (Immanuel Kant[2])

> There are broadsides from those who believe they can see, or even prove, that freedom is inconsistent with the assumption that actions are causally determined, at least if the causes can be traced back to events outside the agent. ... I know of none that is more than superficially plausible. Hobbes, Locke, Hume, Moore, Schlick, Ayer, Stevenson, and a host of others have done what can be done, or ought ever to have been needed, to remove the confusions that can make determinism seem to frustrate freedom. (Donald Davidson[3])

> I believe a '"can" of freedom' which holds in the face of physical impossibility is pure nonsense. (G. E. M. Anscombe[4])

In the two centuries that separate the first and second of the above quotations from the third and fourth, there has been no let-up in the entrenched opposition about the problem of liberty and necessity. Philosophy's best and brightest are, it seems, as uncomprehending of the opposing understanding of free will as they were two hundred years ago; as they were a hundred years before that.[5]

Perennial philosophical oppositions are, as such, unremarkable and commonplace. Philosophers are as divided as ever they were about truth, the mind/body problem, how to respond to scepticism about knowledge, justice and inequality, the *summum bonum*, the cognitive status of moral judgements and, not least, on the nature and purpose of philosophy itself. Controversy and disputation are the very stuff of philosophy. So why should we expect that matters be any different with the problem of free will?

The attitudes expressed in the above quotations, however, reveal something deeper and more unyielding than we are apt to find in disputes about the nature of mind, language or knowledge. They do not register mere disagreement with philosophical opponents, but rather an attitude that the

opposing view is hopelessly muddled or worse. Of the four philosophers quoted, Hume takes the least pejorative attitude to those who do not share his views, but even here the implication is clear: those who say otherwise are refusing to face what are 'plain and acknowledged matters of fact'. The tone of each of the quotations is such as to imply that the opposing party is wantonly ignoring what any honest person can see to be as plain as day. Such claims are all the more striking when one thinks of the persistent nature of the free-will problem and the Byzantine character of some of the reasoning with which it has been argued out. It is one of philosophy's toughest problems - some say ***the*** toughest - and its unyielding resistance to resolution, despite the high quality of some of the attempts to resolve it, would suggest that whatever it is that is being argued about simply cannot be as blindingly obvious as our four thinkers (and many others) have taken it to be.

To this we may add that a real debate about free will has taken place over the centuries, which has not been an endless rehearsal of old arguments and which has involved genuine, detailed, meticulous and substantial argument over new terrain. The contemporary debate about free will has raised new issues and dealt with old ones in new ways. Contributions such as Peter Strawson's treatment of the justifiability of moral responsibility and the reactive attitudes,6 Frankfurt's hierarchical account of free will,7 and the increasing complexity of indeterminist theories of free action have all served to increase the sophistication and clarity of the free-will argument. But beneath the sophistication, painstaking argument and occasional willingness to grant aspects of the opposing case are rooted and seemingly immovable party commitments, which can surface from time to time in vituperative attacks on integrity and character.[8] The opposing party may, it will be granted, advance some subtle and ingenious arguments, and claim a battle or two, but it doesn't alter the fact that for all their cleverness the other lot are at best confused and at worst dishonest or self-deceiving: 'After all,' it might be said, 'who but the morally or intellectually muddled would want to live in a clockwork universe' or 'a world of absolute responsibility and desert.' The arguments can be dazzling and recherche, but the commitments and the reasons for making them are, so we should think, simple and straightforward.

Why, then, are the commitments of the opposing parties so entrenched and unyielding? Is it that, as each side suspects, one of the opposing camps is lacking in some deep and important way which leads them to deny what is undeniable? It would be bad news for philosophy if that were so, since its greatest exemplars are to be found on either side of the divide. Or is it instead that the two sides have, down the long years, been arguing

at cross-purposes? Given the diversity and vigour of the several attempts to lay the problem(s) to rest, it would be strange if this had not been suggested before. And, of course, it has been.[9] Why then is the division so immovable and how can we get beyond it?

MEANINGS, ATTITUDES AND ILLUSIONS

According to several recent treatments of the problem, the argument has not been resolved because it is founded on certain false assumptions shared by both compatibilists and incompatibilists, most notably that the key terms in the dispute have settled and singular or central meanings.[10] The dispute is irresolvable on its own terms because there is no unified meaning of ***free will*** and cognate terms.[11] If we say that Sheila acted freely when she poisoned her neighbour's cat, we will not have referred unambiguously to natural (or even supernatural) properties that are thought to be present in what she did, because we have no clear and unambiguous understanding of what it is to act freely and responsibly. What we have instead, according to Ted Honderich, are two families of conflicting feelings or attitudes about what is required for agents and their actions to be free and responsible, and desires that they should meet those requirements.12 Galen Strawson makes a similar claim about the groundedness of the traditional dispute in opposing attitudes and experiences we all share. He argues that we are all both natural compatibilists and natural incompatibilists. Natural compatibilism derives, ***inter alia,*** from a general acceptance of determination by heredity and environment, from the fact that we do not choose what we believe, and from a sense of detachment from certain unchosen desires.[13] Our natural incompatibilism issues from our sense of autonomous self-control and, above all, from our experiences of being able to choose.14

An immediate objection to such claims is that they come to no more than the argument that the traditional disputants have been talking at cross-purposes, and if this were true, the way forward would be clear enough: we should all simply say in future precisely what we each mean by the phrases 'she could have done otherwise', 'she chose to do it', 'he acted freely' and so forth. It could easily then be settled whether the freedom implied in each case is compatible with the causal determination of choices and actions. But this, by and large, is what compatibilists and incompatibilists have always done; conceding grudgingly that their opponents do have ***some*** idea of what they are arguing for, but that it is a freedom 'not worth wanting'.15 The claim that the dispute is irresolvable on its own

terms, however, grants that in one important sense the opposing parties have not been talking at cross-purposes; that what they have been arguing about is whether we are free in the sense that matters most to us, and about what (if any) that sense is: which is to say, the sense we have of ourselves as freely acting and deciding, and which warrants ascriptions of moral responsibility and practices and attitudes such as blaming, resentment and gratitude. What is irresolvable, inescapably important though it is to us, is the problem of what is required by a singular sense of freedom that, among other things, is thought to entail moral responsibility.

The charge that the argument has been conducted at cross-purposes might also imply that each side ought to have seen what it was that kept their opponents so resolutely on the other side of the fence. With the claim that our notions of free will and moral responsibility express conflicting attitudes, images and exemplars, we have a more convincing explanation of the failure to grasp the other side's position than one of confusion, dishonesty or psychological lack. The reason the two camps have been so convinced of the correctness of their own views, and hence of the confusion of their opponents, is that they have found support for those views in (some of) their own attitudes about the initiation of actions that can truly be described as free and responsible. The two camps have both been wrong in their shared claims that there is a single conception of the initiation of actions, of free will and of moral responsibility, as well as a single definition of the word ***free*** or a single shared belief about what is required for moral responsibility.[16] But both camps have been right to the extent that their respective conceptions of free will reflect deeply entrenched attitudes we all share. The compatibilist and incompatibilist conceptions of free will have therefore been attentive to different features of a common set of opposing attitudes.

The explanation of inability to comprehend the opposing position can be further elaborated if we think about the effects of philosophical reflection. A philosopher who becomes attuned to one kind of response she has to the idea of determinism, arising from one set of deeply held attitudes about free will and responsibility, will tend to reinforce her responsiveness to those attitudes as she begins to think through and flesh out a philosophical characterisation of her responses and to develop arguments in support of them. She could explain the pull of her opposing attitudes as resulting from a residual failure to fully grasp and absorb the philosophical commitment she has adopted. It has also been suggested, unsurprisingly, that psychological make-up and personal preoccupations could predispose one to be more initially attentive to one set of attitudes, and thereafter to convert wholeheartedly to the philosophical camp that best expresses them.[17]

It is also noticeable that the division about free will roughly parallels a difference of philosophical temperament between those philosophers who look to science for both inspiration and method in philosophy and those (many of them inspired by Wittgenstein) whose attitude to science and theory is altogether more cautious or sceptical.[18]

Any hope of attempting to prove either the compatibilist or incompatibilist case by appeal to our prephilosophical attitudes is therefore misconceived. In the first place our prephilosophical attitudes give conflicting signals about the philosophical dispute, in addition to which it is likely that sophisticated philosophical reflection about free will is a journey whose steps cannot be retraced. Once the theoretical frameworks are in place, and intuitions reinforced by reflection and argument, it is implausible to imagine that we can temporarily disengage their influence on our thinking. If it is true that we all have conflicting attitudes about the initiation of actions, then belief that free will and moral responsibility require choices that are not causally determined is not, as compatibilists have supposed, a fanciful product of theoretical imagination, or just an 'intellectualist trinket',[19] but a reflection of deeply held attitudes we all (or most of us) share.[20] We can each be said to have attitudes that involve wanting and hoping for actions to be initiated in a way that is free of causal determination, as well as attitudes that require only that our actions are voluntary and which are at home with the thought that actions are causally determined. And since what we have are conflicting embedded ***attitudes*** (involving desires), rather than settled beliefs, there is no hope that one side of the conflict can be disproved in any standard and non-question-begging sense.

Another way of explaining the failure to reach agreement in the traditional argument about free will is that our prereflective attitudes about free will and moral responsibility are an irreconcilable and unsystematisable mess. According to Richard Double terms like ***free*** and ***morally responsible*** are used in diverse and contradictory ways, reflecting deeply held images and exemplars of what it is to be human, or what it is to be a person, which structure our sense of ourselves as choosers, actors and participants in human practices and relations.[21] There is no single property or group of properties (real or imagined), he argues, that is uniquely picked out by descriptions involving ***free choice.free action, morally responsible*** and so forth. The traditional dispute, it would seem, is not resolvable on its own or on any terms.

However, if we do have conflicting or incoherent attitudes about free will, there remains a question about which attitudes can be satisfied. Those who have thought that the traditional argument is founded in conflicting attitudes agree with compatibilists that incompatibilist conceptions of free

will, control and responsibility cannot be satisfied, either because determinism is probably true,[22] or because incompatibilist conceptions of free will and responsibility are conceptually incoherent. In support of the latter view, Galen Strawson has argued that to be 'truly responsible' for a choice, one must also be responsible for 'how one is, mentally speaking - in certain respects', since how one chooses is necessarily a function of how one is, mentally speaking. But to be responsible for how one is, one must have chosen how one is. To be responsible for that requires that one must be responsible for the choice, including the principles according to which one makes the choice. In order to be responsible for the principles governing one's choice, however, one must have responsibly chosen those. In that case one must also have been responsible for the principles according to which one made that choice.[23] Infinite regress, it seems, cannot be avoided, except by the desperate remedy of declaring the agent's act of choice to be ***causa sui,*** which involves attempting to explain one unsatisfiable notion by means of another.

Another well-known objection to incompatibilist conceptions of free will is that they require of any free choice that it must be the case that given everything as it was at the point of choice, the agent could have chosen differently than she did. There have been several suggestions about how a categorical ability to choose or do otherwise might be possible, involving indeterminacy at the moment of choice or at some time before it.[24] According to compatibilist critics, the principal difficulty facing any such theory is how the randomness involved in any choice of which this is true could be either rational or under the control of an agent.[25]

If the argument that the old dispute about free will and determinism is founded on conflicting attitudes goes with a claim that the desiderata of incompatibilism are unachievable, it is unlikely to prove attractive to 'libertarian' incompatibilists who believe that we are free and responsible in the incompatibilist sense and whose positive programme is to show how this can be so. But even if the argument fails to bring the traditional dispute to a conclusion, as a new approach it has the considerable attraction of giving a reasonable and believable explanation of the persistence and intractability of the problem, which does not rely on ascribing bad motives or muddle-headedness to one or other of the traditional opponents. On Ted Honderich's account, moreover, we have a metaphilosophical way out of the impasse that could prove attractive to at least some of the erstwhile disputants. His suggestion is that although the truth of determinism does not, ***contra*** compatibilism, leave our moral practices, affective relations and life hopes untouched, we can learn to value those feelings, attitudes, hopes and expectations that are compatible with determinism and to eschew those that must be disappointed by it.[26]

At any rate, if we accept the argument about conflicting attitudes as it stands, we must also accept that there are important senses of freedom and moral responsibility we take ourselves to possess, which in fact we do not. In this respect, those who argue that the traditional debate is irresolvable on its own terms clearly agree with incompatibilist determinists, up to a point, about how things are, and with libertarians, up to a point, about how things would be if determinism were true. It seems that we labour under illusions about ourselves and others, and about actions, not just in beliefs but in deeper and rooted feelings and attitudes. We all (or nearly all) wish to be free in what we do, in a way that cannot be satisfied.

Should we accept this? Do we, that is, lack a freedom we unreflectively believe ourselves to possess, and which is at the heart of many of our most important feelings and attitudes? It is difficult to resist the conclusion that there are no single and settled meanings of terms such as ***free will*** and ***moral responsibility,*** nor any single sense or usage of them - compatibilist or incompatibilist - that is the 'everyday sense', or the sense that really matters to us, which would enable the philosophical problems to be resolved by methods associated with ordinary-language philosophy or by any analysis of the conditions required for the key terms to be true.[27] The conclusion is difficult to resist, in the first place, because with a little attentiveness and reflection about our responses to the idea of determinism in certain contexts, we can see clearly enough that we do have attitudes that seem threatened by it and others that are undisturbed by it.[28]

The traditional antagonists, especially compatibilists, might grant that their reactions to the thought of determinism are of this sort, but argue, as I suggested earlier, that after all this is no more than the residual pull of woolly thinking or lack of perspective. We continue to perceive the sun as moving across the sky, and unsupported objects as borne towards the ground by the force of their own weight, without being led to hypothesise a deep attitudinal resistance to what we know to be the case. It has even been suggested that just as everyday concepts such as hot/cold and heavy/light have been replaced and explained by more precise scientific notions, theorisation 'about the important concept of responsibility should not be stymied by "conflicting intuitions"'.[29] This suggestion is characteristic of the kind of theoreticism to be found in the arguments of many compatibilists and has excluded any possibility of their arriving at a persuasive account of free will. A 'technically precise' analysis of moral responsibility would be a bloodless counterfeit. If the problems of free will and responsibility could be resolved by replacing our woolly everyday notions with well behaved and precise theoretical concepts, they would be philosophical small fry.

The second reason for the persuasiveness of the argument that there are

no single and settled meanings of the senses of *free will* and *moral responsibility* that matter to most of us has already been given: its ability to explain the persistent and entrenched character of the traditional dispute. If philosophy's finest are to be found on either side of the argument, and this continues to be the case despite sustained and often ingenuous attempts to resolve the problem, it is unbelievable that the argument is founded on confusion, still less that it exists because of dishonesty or lack of psychological stature. The argument that the dispute is founded on two sets of attitudes we all have, or are capable of having, is simply the best explanation available and to be recommended on that account alone.[30]

Since the claim that the problems of free will, agency and responsibility are founded on attitudes presents a challenge to the widely shared belief that the resolution of the problems should be sought through an examination of the meaning and use of the key terms, it will be instructive to examine two of the contemporary debates about free will in which such a resolution has been attempted. (Those who need no convincing that a resolution of the problems is unlikely to be found through an analysis of meaning and usage may wish to skip the following section and pass directly to the section entitled 'Attitudes'.)

IFS, CANS AND CONSEQUENCES

In a widely discussed chapter of his *Ethics*, G. E. Moore observed that it is a condition of a person's being morally responsible for an action, that he was at the same time free not to have done it; that he 'could have done otherwise'. If determinism is true, then it seems as if, given the antecedent conditions that cause a person to do as he does, he cannot do otherwise. Moore remarks that the meaning of 'I could have done otherwise' is ambiguous, and that it may be that we use it as a short way of saying 'I could, *if* I had chosen; or (to avoid a possible complication) ... I *should, if* I had chosen'.[31] If my choosing to act is a causal antecedent of my acting, then of course *if* I had chosen differently I might have acted differently, and if saying that I could have acted differently means just that I would have done so *if* I had chosen to, then the ability is quite at home with the causal determination of choices. If moral responsibility requires not only that I could have done otherwise than I did but also that I could have chosen so to do, then again this may mean only that I should have chosen if I had chosen so to choose. And again, the truth of such a conditional would be compatible with the truth of determinism.

J. L. Austin cast doubt on Moore's arguments by questioning whether 'could have if I had chosen' is equivalent in meaning to 'should have if I had chosen', whether in either case the *if*s the ***if*** of causal condition, and finally whether the inclusion of ***can*** or ***could have*** in a sentence implies or warrants an *if* clause. The answer in each case, Austin thought, is no.[32] In answer to the second question, Austin observes that if 'I can if I choose' or 'I could have if I had chosen' asserts a causal relation between choice and action, it would follow that **'If** I cannot, I do not choose to' or **'If** I could not have, I had not chosen to', as it is characteristic of any causal conditional that it licenses the inference of its contrapositive. Since, however, 'I can if I choose' implies 'I can, whether I choose to or not' (what Austin describes as the 'all in' sense of can) the *if* in question cannot be the ***if*** of causal condition.

The argument against the conditional analysis of ***can*** has subsequently been taken up by Roderick Chisholm and Keith Lehrer. Chisholm's argument is that 'he would have x-ed if he had so chosen' is compatible with 'he was unable to choose to *x*', that the latter implies that 'he could not have chosen to *x*', in which case 'he would have x-ed if he had so chosen' cannot be equivalent to 'he could have x-ed'.[33] Compatibilists have replied that Chisholm's argument can be circumvented if ***choose*** is replaced by a verb that is not a verb of action, such as wanting or willing.[34]

Keith Lehrer notes that for logical equivalence to be claimed between a statement containing ***can*** or ***could have*** and a conditional statement, it must be logically impossible that the one should be true while the other is false. It is logically possible that ***willing, wanting, choosing*** or any other verb contained in the antecedent of a conditional analysis might also be a ***necessary*** condition for an agent's so acting. It is also logically possible, therefore, that as a result of not so willing an agent is rendered unable so to act. In that case the conditional statement 'S would have done otherwise if he so willed' would be true, although it would be false that 'S could have done otherwise'. The two statements cannot therefore be logically equivalent, and the compatibility of the former with determinism cannot be inferred to the latter.[35]

Lehrer's argument has been criticised on the grounds that it entails that attribution of solubility to a substance cannot mean that it dissolves when placed in water. Being placed in water is a necessary condition of a substance's dissolving, so if a substance is not placed in water it cannot dissolve. Analysis of dispositional powers by causal conditionals, however, has strong intuitive plausibility. The obvious conclusion is that although being placed in water is a necessary condition of a substance dissolving, it cannot be a necessary condition of ***solubility.*** Likewise, it

may be argued that although willing to ***x*** is a necessary condition for ***x-ing,*** it is not a necessary condition for being able to ***x.***[36]

Whatever the merits of the various contributions to the argument about conditional analyses of ***can*** and ***could have,*** there is reason to doubt that a correct analysis of what we mean by them will be given by a causal conditional statement with an appropriate antecedent verb. The concept of ability, expressed by the terms ***can*** and ***could have*** in the senses that are relevant to free will and responsibility, necessarily involves the concept of action. Our concept of action, by and large, is such that the events we call actions have as their initiators (in some unspecified way), agents rather than events. The causation of actions by agents need not be incompatible with their also being caused by antecedent states or events, provided the antecedents, the mechanisms of causation, and their relationship to bodily movements, are of appropriate kinds for us to identify them, without loss, with our prereflective understanding of an agent's relationship to her actions. But that relationship, as ordinarily understood and expressed in action statements, is not one in which actions are immediately caused by events.[37] Since the ***can*** of ability is the ***can*** of ability to act, at least in contexts where freedom and responsibility are at issue, it follows that what is meant by statements of ability is likewise the abilities of agents to act. Statements of ability, like statements of action, are therefore silent about whether specific causal conditions might be sufficient for our being able to act.[38]

As Donald Davidson has argued, therefore, the problem of whether a satisfactory conditional analysis of ***can*** and ***could have*** is possible is bound up with the problem of whether an adequate causal analysis of action is possible.[39] In line with this, although the meaning of ***can*** and ***could have*** cannot be explicated in terms of causal conditionals, it remains possible that causal conditionals could provide the truth conditions of ability statements. This would not satisfy incompatibilists, however, who would rightly note that to give the truth conditions of ***can*** and ***could have*** statements in this way would fail to distinguish them from statements involving ***can*** and ***could have*** that describe the causal powers of inanimate objects.

Even if advocates of the conditional analysis were to succeed in making a convincing case for the possibility that statements of ability are shorthand for causal conditionals, Peter van Inwagen's 'Consequence argument'40 gives independent and intuitively compelling reasons for doubting that what we mean by 'she could have done otherwise' is compatible with determinism. The argument comes roughly to this: if, as is implied by determinism, all physical events are consequences of laws of nature and events in the past, then since we have no control over the laws of nature and what is past we cannot have control over their consequences, includ-

ing the movements of our bodies. The details of the argument can be summarised as follows. Supposing an agent *A* has the opportunity to raise his right hand at time *T*, but does not, then if determinism is true, the proposition (H) that *A* does not raise his hand at *T* will be entailed by a proposition (P) expressing the state of the world at some earlier time, together with a proposition (L) expressing the laws of nature. Taking this as the first premise, the argument can be set out thus:

(1) If determinism is true, then the conjunction of ***L*** ***and P*** entails ***H***
(2) If ***A*** had raised his hand at ***T***, he would have rendered ***H*** false.
(3) If the conjunction of ***L*** and ***P*** entails ***H***, and if ***A*** could have rendered ***H*** false, ***A*** could have rendered the conjunction of ***L*** and ***P*** false.
(4) If ***A*** could have rendered the conjunction of ***L*** and ***P*** false, then either ***A*** could have rendered ***L*** false or ***A*** could have rendered ***P*** false.
(5) ***A*** could not have rendered either ***L*** or ***P*** false.
(6) Therefore, if determinism is true, ***A*** could not have raised his hand at T.[41]

A standard criticism of incompatibilist claims that determinism places us under the control of the past, is that they involve a fallacious derivation of the necessity of actions from the necessity of the conditionals connecting them to their antecedents taken together with the *truth* of those antecedents.[42] If the occurrence of the antecedents was not itself necessary, then compatibilists can argue that no necessity attaches to any consequences they entail: even consequences they necessarily entail. In van Inwagen's argument, by contrast, propositions expressing the antecedent causes of actions are treated as necessary (since they refer to what is past and therefore beyond our control), in which case the inferred necessity of the consequences follows from both the necessity of the conditionals connecting antecedents and consequences and the necessity of the antecedents.[43]

Compatibilist counterarguments have focused on the third premise of the argument set out above, which claims that if what one does is entailed by ***L*** ***and P***, then one is only able to do otherwise than one does if one is also able to render the conjunction of ***L*** and ***P*** false. According to David Lewis, this can be interpreted as making either a weak or a strong claim about counterfactual ability.[44] The weak claim is that for one to do otherwise than one does requires that if one were to do it, the conjunction of ***L*** and ***P*** would be falsified, i.e. that some law of nature or some statement

of an actual past state of affairs would have been false if one had done otherwise.[45] The strong claim is that for one to do otherwise than one does requires that if one were to do it, one could break a law of nature or alter the past, or cause an event that would break a law of nature or alter the past. The strong claim is incredible, but according to Lewis it is only the weak claim that follows from determinism. Lewis's argument for this is that if *A* were to raise his hand and thereby to falsify the conjunction of *L* and *P,* his doing so could have been caused by a divergence from the actual course of events - a 'divergence miracle' - occurring some time before he raises his hand and not caused by his raising it.[46] It would be no requirement of *A*'s counterfactual ability to raise his hand, therefore, that he had an incredible power directly to render propositions about the laws of nature or the past false, even though his doing so would mean that such a proposition had been rendered false, and would indeed require it.

It might seem as if dependence on an earlier miraculous event would put the agent in no better position, in relation to the ability to do otherwise, than one that depends on his having the miraculous power. But if universal determinism is assumed to be true, the occurrence of a divergence miracle would be required for the occurrence of any state of affairs described by a counterfactual, without our being inclined to doubt the truth of all counterfactuals. By undermining the claim that determinism has the consequence that the ability to do otherwise calls for a miraculous power on the part of the agent, Lewis's argument puts attributions of counterfactual ability to agents on a par with counterfactuals about natural objects.

The logical status of the inference from the double necessity of statements of past events and laws of nature to the necessity of statements about actions has also been subject to criticism, despite its avoiding the fallacious inference of necessity involved in traditional incompatibilist arguments. Michael Slote argues that consequence arguments assume or depend on, what he describes as the ***main modal principle:*** *(NP · N(P* :) Q)):) *NQ.* According to Slote this principle involves tacit inferences from *'NP'* and *'N(P* :) *Q)'* to *'N(P · P* :) *Q)',* and from the latter to *'NQ',* which assume, respectively, that the necessity operator has the properties of agglomerativity and closure under logical implication or entailment.[47] Slote's argument against the assumptions depends in the first place on finding examples of modal inference that do not obey the main modal principle.

One such example is as follows. If Jules and Jim chance to run into each other at the bank, their appearances at the bank, taken individually, are not accidents, but that the two of them arrive at the same time is. According to Slote, accidentality also lacks the property of being closed under entailment:

it might be an accident that I am alive right now, but no accident that I am where I am right now, even though my being where I am right now entails my continuing to be alive until now.

An obvious rejoinder to Slote's arguments is that judgements of accidentality are not categorical, but relative to background, standpoint or context, and that this must be taken into account in determining their implications, if any, for the scope of agglomerativity and closure under entailment. Slote's response to this is that if claims of accidentality are relative to standpoint or context, this supports the case he is making. If Jules judges that his being at the bank is non-accidental relative to his own point of view, then, supposing agglomerativity to be true for accidentality, he should take the view either that Jim's being at the bank is accidental or that their meeting is non-accidental; whereas he would naturally do neither. Slote's argument, however, is open to the objection that Jules judges that Jim is at the bank non-accidentally ***relative to Jim's point of view,*** and that their arriving at the bank at the same time was accidental ***relative to either point of view.*** But if we combine the contents of the two points of view (their earlier plans and experiences) into some larger encompassing point of view, as agglomerativity would require us to do, then relative to what is in their combined points of view the meeting is no accident and agglomerativity holds.

Slote argues that one reason we would consider the meeting to be accidental, in contrast to the fact of either Jim or Jules being at the bank, is that while the facts of each being there were called for by earlier plans, this was not true of their meeting. We distinguish between accidentality and non-accidentality, however, according to what is or can be foreseen, rather than just what is planned or purposed. Someone knowing of Jules's and Jim's earlier circumstances and plans could in principle have foreseen their meeting, and even have made plans that depended on it. For such an observer, having a viewpoint whose contents encompass those of both Jules and Jim, their meeting would have been no accident. Such an observer might see it as a happy accident that the circumstances and plans of Jules and Jim were such as to cause them to be at the bank at the same time, but knowing of those circumstances and plans she would not be inclined to see the meeting itself as an accident.[48]

The same objection can be pressed against Slote's argument for failure of closure under entailment in respect of non-accidentality. If it is no accident that I am here now, but it is an accident that I am alive right now, the judgements are made relative to different backgrounds: as Slote himself presents the argument, my being here now is no accident because I planned to be here, whereas my being alive right now is an accident

because only a lucky and unintentional swerve saved me from being flattened earlier by a runaway truck. Relative to my plans, however, it is no accident that I am still alive because they did not include my untimely demise.

Slote goes on to suggest that statements about the necessity of the past and of laws of nature are selective in a way that would cancel or restrict agglomerativity and closure under logical implication or entailment:

> When we say of any past event that we can ***now*** do nothing about it, I think we are saying that our ***present*** desires, abilities, beliefs, character, etc., are no part of the explanation of it. And, more generally, the particular kind of factor in relation to which unavoidability exists at any given time, the factor 'selected' by such necessity, is, simply, some factor (or set of factors) that brings about the unavoidable thing ***without making use of (an explanatory chain that includes) the desires, etc., the agent has around that time.***[49]

To suggest that when we say or think we cannot change what has already happened, we are really speaking or having thoughts of desires, abilities, beliefs, characters, etc. is implausible enough, but to say that thoughts and claims about our inability to ***affect*** the past are really about ***explanations*** of it beggars belief. When we say that nothing can change what has happened, we mean by this that ***nothing, categorically, can change what has happened;*** there is no reason to believe that anything more than this is intended or implied.[50] Neither, therefore, do we have any reason to accept Slote's claim that 'the ordinary notions of avoidability, inevitability, and the like involve the idea of being determined in a particular sort of way', which does not carry over to events that are caused by desires, beliefs, character, etc.[51] There is, I conclude, a sense in which our actions can properly be described as necessary and unavoidable which is entailed by determinism. Whether that sense is incompatible with free will or moral responsibility is a separate matter.

What the debate about the conditional analysis of ***can*** and the Consequence argument have in common is an attempt to definitely establish the meaning or proper use of certain key terms ***(can, could have, avoidable, necessary,*** etc.), thereby to resolve the problem of whether the relative and important senses of ***free will*** and ***moral responsibility*** are compatible with the causal determination of actions. Attempting to resolve philosophical problems by focusing on how we use or how we should analyse certain key terms has been the defining approach of analytic philosophy. If little else has been settled, it is that the important singular uses and meanings

of the key terms have not been definitely established, and that there is by now little hope that they ever will be.

Even Moore thought that it was only possible that all we mean by ***can*** and ***could have*** is given by 'would have, *if* he had . . .'. What more could be done to show not only that it is possible that this is what we mean, but that it is in fact what we mean? Compatibilists have no doubt though that logical equivalence can be proved by showing that it is impossible that an appropriately worded conditional analysis should have a different truth value from a ***can*** statement. This, of course, cannot be proved to the satisfaction of incompatibilists, since in an indeterministic world there would be situations in respect of which ***can*** and ***could have*** statements would be true, but causal conditionals would be either false (in respect of past states of affairs) or of indeterminate truth values (in respect of future states of affairs), and in a deterministic world there would be situations in which the conditional statements would be true although ***can*** and ***could have*** statements would be false. [2]

Thus all compatibilist arguments could ever have shown is that appropriately worded causal conditionals are sufficient for attributions of ability at the level of appearances; in other words, that even if we are all deep down incompatibilists, our experiences of actions that satisfy appropriately worded conditional analyses would be indistinguishable for us from experiences of behaviour in which the agent could have done otherwise. Even supposing that such indistinguishable experiences include all possible experiences, and not merely superficial appearances, incompatibilists would still be inclined to regard them as fakesY

By contrast, the argument that no conditional analysis can be equivalent to an ability statement was also shown to fail, unless one is prepared to reject conditional analyses of dispositions and hence to accept the consequence that only objects that are placed in water are soluble. The meanings and uses of the key terms failed to settle the argument about the consequences of determinism for free will and moral responsibility. The claim that the key terms do not have single or unified meanings that would enable a resolution of the argument would appear to be vindicated in respect of the debate about conditional analyses of ability.

I have also argued, however, that statements attributing ability presuppose or involve the concept of action, and that our usage and understanding of ***action*** are such that we attribute actions to agents rather than causation by antecedent events. I suggested, in addition, that what we mean by ***actions*** and their relations to agents implies nothing about causal relationships between actions and antecedent mental states. This raises the possibility that key terms in the debate about free will and determinism are

not systematically ambiguous in relation to the initiation of actions and its compatibility with determinism,[54] but simply *silent* about it.

Like the 'ifs and cans' argument, the debate over the Consequence argument has also been characterised by stalemate. Compatibilists have traditionally stressed the distinction between causes external to the person and those that are internal, and have maintained that if determinism entails that our actions are necessary in any sense, it is a necessity in respect of internal causes and therefore not threatening to free will and moral responsibility. Van Inwagen's argument was intended to show that the necessity that would attach to actions as a consequence of determinism is not merely necessity in a 'relative' sense (and therefore to be bracketed off as irrelevant to considerations about free will and moral responsibility), and that it implies categorically that we cannot do otherwise than we do. Lewis's argument demonstrates that the ability to do otherwise in a determined world would not require a magical ability to falsify laws of nature or propositions about the past. Therefore, the incompatibilist attempt to prove that the necessity entailed by determinism clearly conflicts with our being able to act other than we do, in the sense required for free will and moral responsibility, is unsuccessful.

On the other hand, Slate's attempt to show that the sense of necessity implied by determinism is definitely irrelevant to free will and moral responsibility was also unsuccessful. There is no reason to suppose that agglomerativity and closure under logical implication or entailment do not apply to the necessity that attaches to propositions about the past and about laws of nature, and therefore no reason to suppose that the main modal principle of the Consequence argument is fallacious. If anyone remains convinced that there must be restrictions on agglomerativity and closure under implication such as Slote has argued for, they should ask themselves what practical distinctions such restrictions would enable us to make in ordinary discourse. In other words, why should we expect there to be rules about inferring necessity that would have no use except in resolving a philosophical dispute? Why and how would such rules have evolved? It seems perfectly sensible, therefore, to say that determinism has the consequence that we cannot do otherwise than we do.

The question that remains, and has not been resolved by either of the two central debates about meaning and use, is whether the inability to do other than we do that is entailed by determinism implies that we are not free or responsible in what we do. It will hardly be surprising that an argument about the status of modal inferences (even if it had been correct) failed to resolve the old disagreement about free will and determinism. We might have hoped for better from careful examination of the meaning and

use of key terms, but that has proved to be as inconclusive as the traditional dispute. It appears therefore that the sources of the dispute do operate at a level other than that of the terms in which it is expressed.[55]

Ought we, then, to attempt to come to the kind of recognition Ted Honderich has recommended about the unsatisfiability of deep hopes and desires we all share about the initiation of actions, and thus to an accommodation with determinism (at least in respect of the initiation of actions)? Or must we instead face up to an incessant and inescapable conflict of intuitions about free will and moral responsibility as just part and parcel of the human condition (or, a little more promisingly, the modem Western condition)?

ATTITUDES

The shift in focus from beliefs to attitudes is an important development in contemporary discussions of free will and responsibility and is clearly central to claims that the traditional argument is irresolvable on its own terms. Some consideration is therefore in order about what kind of change in the terms of debate the emphasis on attitudes involves, and what conclusions might be entailed by it.

The locus classicus for the shift from beliefs to attitudes as the means to an understanding of free will and moral responsibility is Peter Strawson's 'Freedom and Resentment'.[56] In a subtle and sophisticated examination of the consequences of determinism for moral responsibility, Strawson suggests that new light will be thrown on the subject if we consider first what consequences determinism might have for the non-moral attitudes we have that bear a certain kinship with moral attitudes. Our moral attitudes, he argues, are the impersonal or generalised analogues of more personal attitudes such as resentment and hurt feelings, and also of gratitude, love (of certain sorts) and forgiveness. The whole range of these attitudes is based on expectations of regard and goodwill, of and towards both ourselves and others. Strawson contrasts these attitudes with the 'objective attitude', in which our interpersonal expectations are withdrawn. The two sorts of attitudes are not mutually exclusive, but they are opposed. To take the objective attitude is to put aside - perhaps only temporarily or partially - reactive attitudes like resentment and indignation, and the expectations of goodwill and regard on which they are based. To the extent one takes the objective attitude, therefore, one forswears or discounts resentment, indignation, blame and praise. There are several reasons for adopting the objective attitude and several ways of doing so: all of which involve some

suspension of reactive attitudes. We tend to feel the objective attitude to be most appropriate, however, towards those who are psychologically abnormal, such as the insane, or morally underdeveloped, such as small children: those who are incapable of reciprocating appropriately to the reactive attitudes.[57] We may also temporarily withhold the reactive attitudes when an agent has brought something about unintentionally or unknowingly. Resentment and blame may be inappropriate when something injurious or hurtful has been done accidentally, rather than because of lack of goodwill or regard.[58]

Strawson points out that it is not thought to be a consequence of determinism that everything anybody does must be accidental or unintentional or produced by psychologically abnormal or morally underdeveloped characters. We have criteria about what conditions warrant the suspension of the moral and interpersonal reactive attitudes, and acceptance of determinism would not entail that those conditions apply to all actions. But even if it did, this should not lead us to drop the reactive attitudes on grounds of their lack of rational justification. Taken as a whole, the reactive attitudes are part of 'the general framework of human life',[59] and are not the sort of thing that can be given an overall justification or stand in need of one. It is simply not open to us to convert wholesale to the objective attitude. And finally, even if it were open to us, since the reactive attitudes are so pervasive and valuable a feature of the range of human relationships, our first consideration in deciding whether to persevere with them would have to be the likely effects for our lives of relinquishing them.

What is it, then, about the shift in focus from beliefs to attitudes in Strawson's arguments that changes the nature of the debate? It is, first, that the reactive attitudes, and related practices such as punishment, are natural human responses to the behaviour of others and that as such they do not call for a justification. Strawson does not, of course, claim that questions of justification for praising, blaming or punishing, and so forth never arise in particular cases. I may be unjustified in blaming you for what you did, if you didn't really mean to do it (say because you had no idea your remarks could be interpreted in that way). Strawson's claim is that punishing, resenting, praising, blaming, etc. do not, as attitudes and practices, require any ***general justification.*** (It is important to keep this in mind, given the tendency of some philosophers to see his pointing out the consequences of giving up the attitudes as some kind of justification for keeping them.[60] I think Strawson's point is rather that to ask for a justification of the attitudes is to assume that we could give them up if we should discover them to be unjustified, or that it could make sense for us to do so, both of which suggestions are absurd.) The second consequence

of the shift in emphasis from beliefs to attitudes supplements the first: it is that there is nothing, over and above their normal conditions of applicability, that could justify the attitudes. If blaming people and holding them responsible were beliefs or judgements, it would always be possible to ask what conditions are required for such beliefs or judgements to be true or warranted and to press the inquiry to the point where determinism becomes an issue. Attitudes, by contrast, are neither true nor false, and are not warranted by anything over and above their standard conditions of applicability. The shift from beliefs to attitudes as the focus of discussion therefore undermines the characteristic assumption of the traditional antagonists that the problem is about what justifies us in treating people as responsible for their actions.

The emphasis on attitudes is shared by Ted Honderich, who, as we have seen, argues that we each have two families of attitudes about the initiation of actions in relation to moral responsibility, moral worth, our futures, and so on. An attitude, on Honderich's view, is 'an evaluative and feelingful thought bound up with desires'.[61] Each attitude has within it certain feelings and desires, in addition to the grounds of those desires. One family of attitudes has, as part of its grounds, an idea or picture of an 'originator' (that which causes actions while not itself being either caused or random), while the other does not. The relationship between this idea and the feelings and desires it grounds is not one of logical connection, but rather one in which we ***regard*** the ground as a reason for the feelings and desires: it is just the way we are.[62] Since the ground is incompatible with determinism, the attitude that contains it is liable to produce the response of dismay (itself an attitude) about determinism.

On Honderich's view, therefore, Strawson's mistake is in recognising only one set of reactive attitudes, whose grounds can be collectively expressed in the requirement that the objects of the attitudes must have acted voluntarily. Honderich agrees with Strawson that there is no question of logically justifying attitudes, but the reactive attitudes whose grounds include origination as well as voluntariness are inconsistent with determinism, since they include a metaphysical ground that is incompatible with it. The issue is not whether the attitudes can be justified, but whether we can persist in them given the likely truth of determinism.[63] The originative attitudes will be displaced, Honderich thinks, by a growing acceptance of determinism. The claim that there are not one but two sets of attitudes, both of which are valuable to us and deeply entrenched in what we want and how we think and behave, also enables Honderich to reject the traditional assumption that we have univocal conceptions of moral responsibility and free will that are written into our language. Terms having to do

with the initiation of actions have a systematic ambiguity that is owed to the two families of attitudes we all share: 'To ***stick with the words,*** in a certain sense, is to fail to get in touch with the reality, the reality of the consequences of determinism.'[64]

DO THE PROBLEMS OF FREE WILL ALL HAVE TO DO WITH ATTITUDES?

The shift to attitudes as the main focus of attention has therefore produced several criticisms of the traditional debate about free will and moral responsibility. The following seem to me to be true and important: (1) that the real issue is not one of whether praising, blaming, punishing and so on can be justified; (2) that there is nothing, over and above their standard conditions of applicability, that could justify them; and (3) that the problems of free will and moral responsibility cannot be resolved by focusing on the meaning and use of words and expressions: the sources of the problem lie elsewhere.

At the same time, certain other claims and assumptions made by Honderich about the content and importance of attitudes are open to question, two of which stand out: (1) that some of our attitudes have a metaphysical content that is incompatible with determinism; and (2) that the sources of the problem are primarily to do with attitudes. One consideration that seems to support the first claim is this: that when we think of a person's actions as being open to an explanation in terms of their causal history, as determinism implies, it seems difficult to persist in certain feelings of vengefulness, resentment, or even gratitude. The same loss of confidence in these attitudes can be brought about by thinking of a person as a collection of dispositions or states. The natural conclusion is either that such attitudes are inconsistent with determinism or that their grounds include indeterministic propositions or images.[65]

A different way of viewing the matter, drawing on Peter Strawson's discussion of the reactive attitudes, would be that to think of a person's actions as determined is to adopt the objective attitude towards her: an attitude that involves viewing her as something other than a person. To view someone as a person is precisely to feel her to be an appropriate object of the reactive attitudes and to have those attitudes in respect of her. To take an attitude to someone that excludes viewing her as a person necessarily limits our having or being able to have reactive attitudes in respect of her. To think of someone as determined is to adopt one kind of objective attitude towards her, but to adopt the objective attitude towards

someone is not necessarily to think of her as determined: a doctor can think objectively of a patient as a healthy or an unhealthy body, without entertaining any ideas about the determination of her mental states or actions; a quantum physicist with libertarian inclinations might take an objective view of a person's decisions and actions as products of a neural system that is indeterministic and unpredictable,[66] and necessarily won't think of them as determined. The loss of confidence in the reactive attitudes that we can feel in response to thoughts of determinism, therefore, comes about because thinking of people as determined is a species of the objective attitude and not because determinism, as such, is incompatible with the reactive attitudes.[67] We may also note that while it is possible to think of persons in this way, it does not follow that this is how we ought always to think of them, and not thinking of them in this way does not imply that we thereby regard them as having special metaphysical powers. It is also worth noting that the initial judgement that our reactive attitudes are threatened by determinism was not arrived at by an examination of those attitudes, but by certain characteristically ***philosophical*** reflections about 'how we should feel if . . . '. A reflective judgement about how our attitudes are affected by the thought of determinism is in no way a proof that the attitudes have a propositional or imagistic content that is related to determinism.

A different reason for supposing that we have attitudes with anti-deterministic metaphysical contents can come to mind in response to thoughts or hopes about the future.[68] Given the choice, most of us would, I think, prefer a genuinely open future with real possibilities to one that has merely the appearance of being open because we do not know what it contains. Such metaphysical preferences can be thought of as related to hopes and desires that we will not be limited by our past failures and weaknesses of character: that we will be able to escape or transcend our past failings.

Our hopes for unfixed futures can be related to something else that has figured significantly in discussions about free will and determinism: our capacity to decide or choose between alternatives. Several philosophers have thought that it is the experience of facing difficult decisions or 'existential choices', in which it seems that we can literally decide either way - that it is just 'up to us' how we should choose - that is it the heart of our strongest sense of being free and responsible.

If it is a consequence of determinism that it is fixed inescapably that I will never overcome my awkwardness and inertia about learning French, then I have a hope that is inconsistent with determinism, or, as we should say, my hope of learning French requires a ground that is inconsistent with

determinism. Likewise, if determinism has the consequence that all my future decisions are already fixed (e.g. that it is true now that I will decide not to challenge a colleague's misuse of power and opt instead for a quiet life), then my decisions will be other than I experience them to be and other than I want them to be.

Does it follow that my hopes about the future and my experiences of making decisions have anti-determinist metaphysical contents? It does so only if it *is* a consequence of determinism that the future is fixed and inescapable. It seems difficult to resist the conclusion that this is a consequence of determinism, and indeed it could be claimed that it is no more than an expression, in terms of the future, of what determinism means. Nevertheless, and despite the counterintuitiveness of saying so, I think there are good reasons to reject the claim that fixed futures are a consequence of determinism. Those reasons are set out in Chapter 4, and without anticipating them let it be noted for now that the claim that our hopes about the future and our experiences of making decisions have anti-deterministic metaphysical contents depends on an arguable (although very plausible) premise.

Let us consider one final reason for supposing that we have anti-deterministic attitudes about the initiation of actions. As I remarked earlier in respect of the causal analyses of ability, our relationship to our actions, as ordinarily understood and expressed in statements about actions, is not one in which actions are directly caused by events or states. Our understandings and statements about actions reflect something else: that we do not ***experience*** our actions as directly caused by events or states. We do, it is true, often experience actions as preceded or driven by states and circumstances, such as intentions or fear, which we may think of as having some form of causal relationship to our actions, but we do not experience ourselves as agents, or our role in relation to our actions, as in any way identical with our intentions or fears. Many incompatibilists have concluded from this that how we understand and experience our actions is incompatible with their being causally necessitated.

Once again, however, it would not follow from the fact that the way we experience and understand ourselves as agents lacks a causal content, that the said experiences and understandings are incompatible with our actions having causes. It may be, as I suggested earlier, that our experiences and understandings of agency are simply silent about causation; and it may even be true that their intelligibility entails or presupposes some causal relationship. None of the examples considered, therefore, of attitudes about the behaviour of others, about life-hopes, about decisions and about actions, forces the conclusion on us that our attitudes and experiences have

anti-deterministic metaphysical contents, or entails, if determinism is true, that we are subject to metaphysical illusion.

It may also have been noticed, in respect of the last two examples given - those having to do with decisions and actions - that where determinism appears to be threatening to free will, it is not only attitudes that are threatened by it but also certain characteristic experiences. The shift in emphasis from beliefs to attitudes is important, but it would be a mistake, having abandoned the notion that problems about free will and moral responsibility can be resolved by linguistic means, to conclude that the sources of the problems are solely or primarily to do with attitudes.

INCOHERENCE

Attitudes are also central to Richard Double's argument that the traditional arguments about free will and moral responsibility are irresolvable. Unlike Honderich, however, Double does northink that we have attitudes that are incompatible with determinism. Since attitudes lack truth values, our attitudes about free will and moral responsibility stand in no definite relation to determinism and do not commit us to any specific responses to it.[69] Our attitudes about free will and moral responsibility are conflicting and incoherent. Any attempt, such as Honderich's, to regiment the attitudes into an orderly system, or to attempt to say what we want by way of free will and responsibility on the basis of them, will be futile.

One reason that Double offers in support of the claim that our attitudes about free will and responsibility are incoherent is, as already mentioned, that the way we are apt to think and talk about free will and responsibility reflects deeply held but contradictory images and exemplars of what it is to be free. Take the definition of acting freely, considered at the beginning of this chapter, as being able to do what you want or most want. This may be said to involve an image of the free person as someone who is able to reflect on and evaluate her competing wants, to determine what she really wants and to act on that. The image is one in which a capacity for rigorous reflection is accompanied by a strong-willed ability to put one's judgements into action. Contrast this now with the image of freedom involved in situations of existential choice, which, as I mentioned earlier, has been thought by several philosophers to be the paradigm context for free will. It is a characteristic of existential choices (such as Sartre's young man who must choose between joining the resistance or remaining with his ailing mother) that reasoning and evaluation give no clear recommendation about how to choose and, indeed, that the lack of any clear recommendation is

what makes such choices so quintessentially free. The contrast between these paradigmatic examples of what it is to be free is no doubt a principal source of the conflicting intuitions that have led to the opposition between Stoic and compatibilist accounts of freedom as the ability to choose and act for the best, or for what we most want, and the incompatibilist emphasis on the 'liberty of indifference' - the capacity categorically to choose either way - as the freedom we want and value. And as they stand, the two paradigms seem to give irreconcilable indications about what we want of free will and responsibility.

It remains possible, however, that the two paradigms[70] attach to different features of our experiences and attitudes about free will, responsibility and agency and that the apparent irreconcilability will be dispelled by showing how they do. (It should also be kept in mind, and without wishing to imply that our many and diverse ways of thinking, talking and acting must all cohere, that if we do use the term ***free*** in fundamentally contradictory and irreconcilable ways there is something of a mystery about how it has come about that we do so. We know, of course, that there are contradictory uses of the term ***freedom*** in politics, but we also have something like an explanation of that. It is doubtful that if the way we talk about ***free will*** and ***responsibility,*** or the attitudes we have in respect of them, are contradictory, we will easily explain the fact in terms of conflicting interests, ideologies, and so forth.)

We therefore have two possibilities: either our attitudes about free will and responsibility are incoherent and irreconcilable, or they can be reconciled by carefully distinguishing and examining the different contexts and experiences to which they attach. The obvious way to decide the matter is by attempting to see if the attitudes can be reconciled in the manner proposed. To object that such a proposal envisions reconciling the unreconcilable or regimenting the unregimentable would simply beg the question.

If we are to distinguish the various contexts and experiences that might give rise to our conflicting intuitions about free will and responsibility, how might we begin? Several distinguishable problems have already been identified in respect of the consequences of determinism and divergencies in our attitudes and intuitions about free will, including the ways in which we understand actions and decisions, hopes about the future, and existential or incommensurable choices. All of these, in turn, can be and have been discussed and examined separately from the large issue that many have taken to be ***the*** problem of free will: the problem of whether determinism is compatible with moral responsibility.

Is this a promising way to proceed? The issue we face is whether the

problems of free will are resolvable. As a strategy for settling this question, and any resolution that might depend on it, distinguishing the compatibility of moral responsibility and determinism from the other problems about free will and agency invites the objection that such a distinction would be a mere philosophical contrivance, and one that is unlikely to result in any conclusion that will prove satisfactory to the many philosophers whose worries about free will ***just are*** worries about moral responsibility and what is required for it. One reply to this is that if we fail to come to any philosophically satisfactory conclusion about moral responsibility, we may still succeed in clearing up the several further problems about free will that can be distinguished from it, and thereby come to a clearer view of what is at issue in the major remaining problem. An entrenched incompatibilist, for example, might find the idea of blaming and punishing determined creatures deeply unfair. She may also believe that to think of people as determined is to imagine them to be no more than calculating mechanisms with no real capacity for acting and deciding. These two responses to the thoughf of determinism may well tend to reinforce each other, with the effect that how plausible she finds arguments in respect of the one idea will be tacitly affected by her attachment to the other.[71] The ideas are of course related, but they are also distinguishable and can and should be examined on their own merits.

A second reply to the objection is that those philosophers who have taken an interest in the problem of free will have not all supposed that there is one unitary sense for each of the key terms. Neither have their preoccupations with the consequences of determinism been of one uniform kind and nor have they had the appearance of being so. In addition to wondering whether our conception of free will and moral responsibility is such as to require a contra-causal ability to do otherwise than we do, philosophers have also been concerned with the consequences of determinism for the kind of freedom we take ourselves to have in acquiring knowledge and determining the truth,[72] for deliberating and deciding and for our understanding of ourselves as agents. In the argument about the compatibility of deliberation with causal determination, for example, few of the protagonists have thought that deliberation has a necessarily moral character. What has been thought of as significant is the process of deliberating about what to do, which is considered to be incompatible with the belief that how one will decide is settled in advance. Deliberation is often concerned with moral matters, but not always.[73] A person might deliberate about a career choice or about the best way in which to make a philosophical argument. Moral decisions, by contrast, can sometimes be straightforward enough to involve little deliberation.

It might be argued that philosophers have been concerned about agency because agency is required for moral responsibility and it is this sense of agency alone with which they are really concerned. The claim is undermined by a philosophical literature directed at analysing a basic concept of agency that extends in scope to non-moral beings such as small children and spiders.[74] Turning instead to a more exclusive idea of agency we are thought to possess as human beings, we may note that there is one important image or idea of fully human agency that is described as involving the ability to stand back and evaluate conflicting motives, and to judge between them in such a way that one is able to intervene effectively in one's behaviour by throwing one's weight behind the motive one has identified oneself with.[75] Thus stated, the description has been taken to conflict with causal determination of decisions and actions by antecedent events; and thus stated, the description is not in any way conceptually dependent on the idea of moral responsibility. Our conception of moral responsibility does require, above all, a moral agent, but our concept of agency is not so obviously in need of the concept of moral responsibility.

A further reason for thinking that the problem of moral responsibility can and ought to be distinguished from other problems of free will comes in the form of an argument against two claims made by Richard Double: that since there is such a close mutual dependency between the terms ***free*** and ***morally responsible,*** the former, like the latter, must be a moral concept, and that the conditions of application of ***free*** and ***morally responsible*** are identical.[76] That both claims are false can be seen from one counter-example: the problem of weak-willed actions. Our common practice is to hold weak-willed agents responsible for their actions. According to the Stoic conception of freedom, a person has free will to the extent that she is able to will and act in accordance with what is right and reasonable. Since weak-willed actions are performed against the agent's better judgement, they must necessarily be judged as unfree according to the Stoic conception. Now one may say 'so much the worse for the Stoic conception of freedom, since it is clearly at variance with what we mean when we say that a person acted freely'.[77] Alternatively one may argue that such a conception of freedom implies a related conception of moral responsibility, or, perhaps, that there is a basic sense of moral responsibility that is 'good enough' for weak-willed agents but falls short of 'true responsibility'. However we respond to the Stoic conception, it remains the case that what it takes to be required for free will can be clearly distinguished from issues about moral responsibility; enough for philosophers who assent to the Stoic conception of freedom to have quite divergent views about what conception of moral responsibility this commits them to.[78] The possibility

of divergent judgements about whether weak-willed agents are morally responsible therefore refutes the argument that considerations about responsibility and freedom are interchangeable.[79]

There is therefore no compelling reason to treat *free will* and related terms as necessarily moral, or as conceptually dependent on moral responsibility, blameworthiness and related notions. And since there are things that matter to us that are bound up with our beliefs and desires about free will, and which can be expressed without any thought to moral responsibility, there is a good case for distinguishing problems of free will that do not imply the concept of moral responsibility, in the hope that in being thus distinguished both sets of problems can be dealt with and resolved on their own terms.[80]

CONCLUSION

We began with a set of worries about whether we are really agents in the unreflective everyday sense we take ourselves to be and, if so, whether we are or can be free and responsible in what we do. The philosophical argument about the answers to those questions is now centuries old, and as yet there is little sign of a resolution that is likely to command common assent. The view of a number of contemporary philosophers is that the argument is unlikely to be resolved on its own terms, since the opposing positions are sustained by entrenched attitudes and images that provoke contradictory or incoherent responses to thoughts about determinism and natural causation. This is thought to explain both the longevity of the traditional dispute and the embattled and uncomprehending responses of each side towards the other. Following Peter Strawson's seminal discussion of moral responsibility and the reactive attitudes, these philosophers have argued that there is no question of providing the kind of justification the traditional antagonists have sought for our practices of holding people responsible and blaming and punishing them. It has also been urged that the traditional attempts to settle the argument by focusing on settled, univocal and prephilosophical meanings of the key terms is misconceived, since what we mean by the key terms is ambiguous as between opposing or incoherent collections of attitudes. The claim that the traditional argument cannot be resolved by a careful examination of meaning and use is supported by the examination of two of the classic debates around which the traditional argument has been conducted in this century, both of which have resulted in stalemate.

Philosophers such as Ted Honderich and Galen Strawson have proposed

a resolution of sorts, involving a coming to terms with the fact that all or most of us have attitudes with metaphysical contents that cannot be satisfied if determinism is true (and, at least in Strawson's view, cannot be satisfied even if it isn't true). Nevertheless, the proposed resolution would not be an entirely unhappy one: we may yet be free and responsible in a recognisable and important sense (the sense that has been set out in increasingly sophisticated detail by compatibilists). The kinds of free will and responsibility sought by incompatibilists are not the only ones that have value for us. Even Richard Double, who does not think any resolution of the dispute is possible, allows that we can still have much that we want by way of free will. We can, despite determinism, plan our futures, pursue goals, imagine alternatives and (sometimes) get what we want. Determinism does not imply fatalism, and there is an intelligible sense in which we can do otherwise than we do and in which we are not the helpless puppets of blind and ineluctable forces.

What we cannot have, seemingly, is the relationship we take ourselves to have to our actions, in which they are initiated by us as agents rather than by dispositional states or causal sequences. Neither can we have the futures we would hope for, in which we are able to overcome those limitations of character that have been the source of our past failures.[81] Nor, it seems, can we have really open choices, such that it is truly 'up to us' what we do.[82] To repeat: none of this is to say that we are left with a lesser freedom or a lesser responsibility. Still, we are left without something we want and value (although, according to Honderich, there is the prospect that we may bring ourselves not to want it). For the time being, at any rate, we want what we cannot have and we are subject to metaphysical illusion.[83]

In considering these arguments, I have suggested that there is room for doubt about whether we are subject to illusion. I have pointed out that the threat that determinism seems to pose to the reactive attitudes, or to our hopes about the future, is arrived at by typically philosophical reflections involving arguable premises. I have also suggested that apparent inconsistencies in images and attitudes we have about what it is to have and exercise free will may be overcome by clarifying the different contexts and experiences to which they attach, and particularly by distinguishing the problem of moral responsibility from various other puzzles about free will and agency. Finally, while accepting that the sources of the problems lie elsewhere than in the meaning and use of the key terms, I have suggested that those sources are not confined to attitudes and feelings but also include our experiences of acting, deciding and choosing: experiences which, like our attitudes, may be rather narrower and more metaphysically innocent

than Honderich, Galen Strawson and many incompatibilists have taken them to be.

I now want to suggest that a further reason why the traditional antagonists are as far from reaching agreement as they have ever been is that just as traditional compatibilism failed to do justice to our unreflective feelings about moral responsibility, blame, gratitude and the like,[84] so also has it failed to get to grips with important features of the phenomenology of acting, deciding and choosing. The modem 'hierarchical' compatibilist accounts of free will of Harry Frankfurt[85] and others, according to which free will consists in an ability to act on one's chosen or embraced desires, are a definite improvement on Hobbes's classical conception of motivation by one's strongest desire. But compatibilist accounts have been strongly 'objectivist', just in the sense that they have been concerned with the capacities (self-motivation, self-consciousness, reason, etc.) that are required for free action and free will, and the structure of mental causation that is required for such capacities. They have had little to say about the phenomenology of freedom, which is to say, the way we experience ourselves as actors and deciders. If I am right in suggesting that many of our beliefs and desires about freedom are substantially owed to our experiences of being able to act and decide (including our understanding of ourselves as the authors of our acts and decisions), this represents a major lacuna in compatibilist arguments. Incompatibilists have rightly perceived the lack in compatibilist accounts of free will and agency and have either concluded that we cannot be free, or attempted to rectify the deficiency with theories of action involving indeterminacy or causation of actions by agents rather than events. Ted Honderich and Galen Strawson concede that what compatibilists have offered is not enough, but argue, nevertheless, that it is the best we can have and is certainly worth having.

In contrast to both the traditional antagonists and their contemporary critics, I want to argue that we can improve on existing compatibilist arguments about the kinds of freedom and agency that are possible; that we can rectify the deficiencies of existing compatibilism without appealing to any theory that places our actions beyond the scope of determinism or natural causation. In the case of moral responsibility and the family of moral and non-moral attitudes and practices to which it belongs, Peter Strawson's arguments have already gone a long way towards rectifying the deficiencies of traditional compatibilism. Nevertheless, there are two reasons for thinking that more can be done. They are first, that despite Strawson's arguments, philosophers are still apt to ask whether our practices of blaming, punishing and holding people responsible for their actions are fair or justified (and, as I said, have even taken Strawson's

arguments as offering a justification); and secondly that Strawson's account lacks a convincing explanation of why philosophers are inclined to think a justification is required.

The remaining areas in which we may seek to remedy the deficiencies of existing compatibilist accounts is in respect of incommensurable choices and a related image of ***ultimate responsibility,***[86] of hopes and beliefs that our futures are not already fixed and inescapable, of what is required for our experiences of deliberating and deciding to be veridical, and, likewise, of what is required for the way we understand and experience ourselves as agents (both in a basic sense we share with animals and in a distinctively human sense in which we take ourselves to have a capacity to stand apart from and evaluate conflicting motives and to act decisively in favour of those motives with which we identify). A further issue whose resolution is bound up with the answers to questions about deliberation and action is whether, when confronting alternative courses of action, we can take ourselves, without illusion, to have reason for what we do and to have control over it whichever course we do take.

It would be heroically optimistic to suppose that the kinds of argument I will make for the veridicality of our experiences and understanding of freedom will lead incompatibilists finally to concede the point. If they have not been won over by an account of moral responsibility that does as much to meet their concerns as Peter Strawson's, what reason is there to suppose that any additional insights I can offer will be more enticing to them? Moreover, as I argued above, it is a recognisable effect of arguing for one's attitudes and intuitions that those attitudes and intuitions tend to be reinforced and hardened. Entrenched resistance to determinism, or to the compatibility of free will and determinism, is unlikely therefore to be broken down by what I will have to say. But if it is possible to show that the experiences and attitudes to which incompatibilism has attempted to give expression do not have contents that are incompatible with determinism, then incompatibilism, like a defeated scientific ontology, might attract a diminishing number of philosophical adherents: the die-hards would die out.

Further reasons for expecting a loss of support for incompatibilism are provided by the failure of contemporary incompatibilist theories to give coherent theoretical expression to what it is they think they want, and by the attitudinal perspective on the traditional dispute. If it is true, as Galen Strawson has suggested, that it is close to impossible for most of us to be genuine incompatibilist determinists, then what alternative is there but to concede that there is a free will worth having that is compatible with determinism, or, as Richard Double has done, to declare the dispute to be radically irresolvable?

The attitudinal perspective on the traditional dispute suggests a promising new tendency of refusing the embattled and unyielding attitudes of the traditional combatants by thinking metaphilosophically about their causes and presuppositions; above all in being prepared to reflect on what there is in each of us that can make either of the traditional positions seem the right one. I will argue that the attitudinal account of what it is that has made incompatibilism seem compelling is mistaken. Nevertheless, attempting to come to an understanding of what it is that has made the opposing positions seem compelling (and what can be done to go beyond them) is in itself an important step in the right direction and offers a real possibility of passing beyond the traditional hostilities.

What incompatibilists have wanted (or attempted to give expression to), and what we should all want, in respect of these issues, is for our unreflective attitudes and experiences not to be illusory. In the chapters that follow I will show that such wants can be satisfied; in other words, that we can blame and resent and feel gratitude, and that we can act and decide - and sometimes do so freely - in the ways we unreflectively take ourselves to do, without our doing so involving us in deep-rooted metaphysical illusions.

2 Moral Responsibility

The argument about whether moral responsibility is compatible with causal determination of actions has focused on the meaning, role and implications of the claim that it is a necessary condition of holding someone responsible for what she does that she could have acted differently. In the first chapter I suggested that two of the twentieth-century arguments about what this condition means and what is required for it to be true have been inconclusive. Compatibilists have also argued that the role of the 'could have done otherwise' condition[1] is in stipulating that a person is to be held to be responsible for her actions if those actions issued from what she wanted or intended, or from her character, rather than from ignorance, interference or unhappy circumstance. According to Frankfurt, it is false that moral responsibility requires that an agent 'could have done otherwise'.[2] If Jones were threatened or in some way coerced we would consider him not to be responsible for what he does only if he acted as he did ***because*** he was coerced. If he were coerced but did what he did because he wanted to, he would still be responsible for what he did. This is true, Frankfurt argues, whatever the sense of 'could have done otherwise'. 'The principle ... should thus be replaced', he suggests, 'by the following principle: a person is not morally responsible for what he has done if he did it only because he could not have done otherwise' and not also because he wanted to do it.[3]

What we are concerned about when holding people morally responsible, according to compatibilists, is whether their actions were the result of ill will or lack of consideration,4 or of morally reprehensible states of mind or attitudes.[5] Whether we possess the ability contra-causally to do otherwise than we do, they argue, is neither here nor there. If a person takes pleasure in the suffering and misfortune of others and deliberately acts on that, then such a one is undoubtedly the proper recipient of the judgement that they are wicked, and of the outrage, condemnation and punishment that follow from that.

Put like this the compatibilist case appears compelling and in tune with our fundamental moral sentiments; and yet it has made little or no impact on the basic incompatibilist conviction that praising, blaming and holding responsible cannot coexist with determinism. What, then, is the source of this basic resistance to the compatibilist case? According to Peter Strawson, compatibilists traditionally have tended to put the case for the compatibility of moral practices with determinism purely in terms of their efficacy

in modifying behaviour, and without regard to the human attitudes and feelings these practices express. Libertarianism is a mistaken attempt to fill in, metaphysically, the missing humanity of the compatibilist account of moral practices. With a proper regard for the full facts of moral and interpersonal practices, no need remains for contra-causal freedom or any other libertarian formulae.

> The optimist's style of overintellectualizing the facts is that of a characteristically incomplete empiricism, a one-eyed utilitarianism. He seeks to find an adequate basis for certain social practices in calculated consequences, and loses sight ... of the human attitudes of which these practices are, in part, an expression. The pessimist does not lose sight of these attitudes, but is unable to accept the fact that it is just these attitudes themselves which fill the gap in the optimist's account. Because of this, he thinks the gap can be filled only if some general metaphysical proposition is repeatedly verified.... Even the moral sceptic is not immune from his own form of the wish to over-intellectualize such notions as those of moral responsibility, guilt, and blame. He sees that the optimist's account is inadequate and the pessimist's libertarian alternative inane; and finds no resource except to declare that the notions in question are inherently confused, that 'blame is metaphysical'. But the metaphysics was in the eye of the metaphysician.[6]

Strawson concedes that libertarian metaphysics may have found its way into how the concept of moral responsibility is commonly understood, in the form of 'reflective accretions',[7] but thinks its incoherence counts against its having any real place in our ***practice*** of holding each other responsible for what we do.

It is now more than thirty years since Strawson's essay first appeared, and there are still a significant number of incompatibilists who remain unpersuaded by it. It would seem, then, as Honderich and others have argued, that libertarian conceptions of moral responsibility run deeper than Strawson's accusations of inane intellectualising suggest.

In the previous chapter I identified two important ways in which Strawson's focus on attitudes changes the nature of the debate about moral responsibility. The first was that since the reactive attitudes and related practices are natural human responses to the behaviour of others they do not stand in need of any justification over and above what can be given by reference to the excusing conditions that are internal to the attitudes and practices. I suggested that Strawson's claims that giving up the attitudes is neither possible nor desirable should not be taken as a putative

justification of them. The point being rather that seeking a justification of the attitudes implies that we either could or should abandon them if they were found to be unjustified, neither of which suggestions can be taken seriously. The idea of seeking a justification for something we know we couldn't and shouldn't stop doing is incoherent. The second and related way in which the focus on attitudes changes the nature of the debate is that there is nothing that could justify the attitudes.[8]

A typical incompatibilist reply to the claim that the attitudes and practices are natural to us, and therefore don't require a justification, is that while we may have no choice about whether to have reactive feelings, we do have choices about whether to ***adopt*** attitudes and whether to engage in the related practices. To blame or punish someone, they will say, is to treat them differently from how we treat others, and differential treatment requires a justification if anything does. And if it is the case that people cannot do otherwise than they do, then pointing out that someone has acted out of malice is insufficient to justify blaming or punishing them.

What should we say to incompatibilist persistence in believing that moral responsibility stands in need of a metaphysical grounding in the ability categorically to act otherwise than one does? I argued in the previous chapter that there is a clear sense in which, if determinism is true, no one can act other than they do, and that the clann that this is so does not run counter to our ordinary uses of ***necessity*** and related concepts. If someone persists in believing that the 'could have done otherwise' condition is a requirement of holding people morally responsible, which must be true in any sense, it is unclear how we can persuade them that this is not so, still less how we can prove them wrong.

This seems to lend support to Ted Honderich's claim that incompatibilism gives expression to one kind of attitude we have in respect of resenting, blaming, punishing and so forth, which is grounded in an idea or image of undetermined (initiation of) action. Can we, then, resist the conclusion either that blaming and punishing do require a justification in terms of a contra-casual ability to do otherwise or that we are constitutionally inclined to think that they do?

One way of viewing the continued incompatibilist commitment to the idea that moral responsibility requires a contra-causal libertarian grounding is to recast Strawson's arguments about the place and status of moral responsibility within the range of reactive attitudes using the Wittgensteinian device of conceptualising forms of discourse and social practices as games. In considering the consequences of determinism for punishment, resentment, responsibility and so forth we are examining the ***morality game.*** The morality game has been substantially modified over the years. Some former

rules and plays that were thought to be inappropriate, or to work badly, were, in time, modified or dropped, just as new features of the game appeared and developed without anyone having planned or intended them.[9] None of this in any way implied that the game itself should be abandoned; quite the reverse, making changes to the game would not make sense without an overall commitment to it. The rules of the game and the way we understand them include criteria for what is to count as obviating responsibility, as well as what can count as desert, or what can count as praiseworthy, etc. There is much, however, about which the rules are silent.

Many games have exact and determinate rules and descriptions that have evolved through practice, argument and discussion. There will have been people for whom the inclusion, exclusion or amendment of a rule by some ruling authority, or a change in some important detail or circumstance of the game, meant that what resulted was no longer authentically the same game. An example of this is provided by those who take the view that growing professionalisation in a sport means that it is being played and determined according to considerations that are strictly external to it.

Incompatibilists are like these disappointed enthusiasts. They see something in the rules about alleviation of responsibility that seems threatened by the change in (perceived) circumstances that would be represented by determinism. Other players and commentators disagree with the interpretation, but the rules and the meanings of the terms they contain are insufficiently specific to decide the matter unarguably. In that case it is not that incompatibilists have a mistaken interpretation of the rules or the game, but that they have a misconceived attitude about what follows if a rule or feature of the game is changed or a decision is taken to interpret it in a particular way. Again, supposing we had a choice about whether to play on or not, continuing to play the game with some feature we don't like is preferable to giving it up altogether; and surely the perceived loss does not make the game entirely unlike what incompatibilists had originally taken it to be. It goes without saying that morality is a feature of life far more important and deeply rooted than any game, and that the criteria for holding people morally responsible are of greater importance than a disputed rule or change in circumstance. Nevertheless, the comparison is instructive. There just isn't enough in our standard conditions for the ascribing of moral responsibility to definitely resolve an argument about whether it is compatible with determinism.

There are circumstances in which it is thought inappropriate to blame, which are detailed by Strawson, but even these are not always clear-cut. Some individuals are more patient and tolerant than others, and more

inclined to try to view wrongdoings objectively than to engage in blaming and punishing; in some cases because they have thought a lot about the causes of wrongdoing. Advising them, when they have chosen not to blame, that it is sometimes appropriate to do so simply misses the point that such individuals often see no value or sense in blaming and punishing. In this they are neither mistaken nor correct. Holding people morally responsible and blaming and punishing them are practices that have evolved over time. There is no reason to expect that the criteria for their proper application will be able to settle a matter their evolution took no account of.

I think this way of viewing incompatibilist insistence on a categorical ability to do otherwise has much to recommend it; particularly in relation to the Consequence argument discussed in the previous chapter. Nevertheless, there is another dimension to incompatibilist claims about moral responsibility that comes to more than an argument about the 'could have done otherwise' condition. Consider the spirit of detachment expressed in the following passage:

> Therefore, on applying my mind to politics, I have resolved to demonstrate by a certain and undoubted course of argument, or to deduce from the very condition of human nature, not what is new and unheard of, but only such things as agree best with practice. And that I might investigate the subject-matter of this science with the same freedom of spirit as we generally use in mathematics, I have laboured carefully, not to mock, lament, or execrate, but to understand human actions; and to this end I have looked upon passions, such as love, hatred, anger, envy, ambition, pity, and the other perturbations of the mind, not in the light of vices of human nature, but as properties just as pertinent to it, as are heat, cold, storm, thunder, and the like to the nature of the atmosphere, which phenomena, though inconvenient, are yet necessary, and have fixed causes, by means of which we endeavour to understand their nature. . . . (Spinoza[10])

Notice that Spinoza does not claim here that the attitude of detachment is justified by an awareness of the causes of human behaviour. What he describes is an endeavour to understand that behaviour: an understanding that must focus on its causes and which requires detachment. This is in line with the suggestion I made in the previous chapter that thinking about the causes of someone's behaviour is one form of what Peter Strawson calls the objective attitude. I also suggested there that the reason for the loss of confidence we may feel about feelings of vengefulness, indignation

and so on, as well as blaming and holding people accountable, is attributable to the fact that taking the objective attitude to someone necessarily involves setting aside the reactive attitudes towards them. According to Jonathan Bennett, since people tend not to think of blameworthiness and accountability in terms of feelings, those who are at all influenced by Spinozism (and, we might add, those libertarians who are repelled by it) are apt to misperceive the loss of their ***feelings*** of indignation etc. as a loss of a judgement that the wrongdoer is to blame.[11] But the loss of reactive feelings, brought on by taking the objective attitude, does not show those feelings or the practices that express them to be inappropriate, and there is no general reason for preferring the objective attitude to the reactive attitude. A Spinozist who strives always to view human actions as strictly necessitated and predictable natural phenomena forgets her own natural needs and inclinations.

As it stands, however, this response is insufficient to meet the Spinozist's concerns. We may allow that it would not be practicable or desirable universally to abandon our practices of holding people responsible or always to adopt the objective attitude towards the endeavours, achievements and misdeeds of others, but this still leaves us with a judgement to make in respect of any particular action. We have already noted that there are those who are less inclined to blame and to resent than others. Their disinclination might be a result of Spinozist reflection on the causes and circumstances of wrongdoing, or a result of innate temperament. Suppose that a person accepts Strawson's claim that there is no reason for, and no possibility of, entirely relinquishing blaming people for their actions, but is sufficiently impressed by considerations about the etiology of human behaviour to doubt whether blame is an appropriate response in many situations (which may be true, in varying degrees, of many of us). Would Strawson and Bennett wish to say that such individuals are guilty of a category mistake? or that the reactive attitudes are always appropriate, providing the criteria for their application are fulfilled? Nothing in Strawson's arguments would support such a strong claim. They show only that there is no reason always to adopt the objective attitude.

It might be argued, since it has been allowed that universal abandonment of moral responsibility and the reactive attitudes has been ruled out on grounds of practicality and desirability, that we should engage in the reactive attitudes and practices whenever the normal excusing conditions do not apply. Such an approach would be contrary to the spirit of Strawson's claims that the reactive attitudes are natural to us and stand in no need of a general rationale or justification. It would conflict, moreover, with what we know about the sorts of conditions that can lead people to refuse to

engage in blaming or lead to qualification or amelioration of their reactive feelings. Someone who knows that the spiteful remarks of her partner were caused by pressure of work can sometimes set aside her feelings of resentment. It does not follow that it would always be right for her to look to the causes of her partner's behaviour or that it would be appropriate for anyone else to do so. Her response will follow partly from her other feelings towards her partner and partly, perhaps, from reflection about how the hurtful remarks relate to his character, how he would subsequently feel about having made them and so on.

Strawsonians could allow that this is so, and that there are indeed a range of cases in which the internal criteria for the reactive attitudes and practices give no unambiguous guidance about how we should respond to bad behaviour, while insisting that there remain many cases where it is clear and uncontroversial that resentment, indignation and blame are in order. Indeed the cases are suggested by what complicates those in which how we should respond is not so clear-cut. A person who is characteristically spiteful, ***rather than reacting badly to pressure,*** for example, is a proper object of blame. One who steals ***without needing to,*** and knowing full well the harm and distress his actions cause to others, gives proper cause for indignation.

Commenting on the Spinozist argument that our proneness to the reactive attitudes is causally dependent on ignorance of the determining causes of our actions, Strawson suggests that it is informed by an absurd vision of human behaviour being brought entirely within the scope of scientific understanding (i.e. as entirely law-governed).[12] But there is a different way of understanding the Spinozist position: one which Strawson (and compatibilists in general) would find harder to resist. Rather than attempting to conceive of human behaviour as entirely law-governed, the Spinozist may see misdeed and malice as caused by misconceptions and limitations of vision. The person who is characteristically spiteful could be understood as having acquired that characteristic as a misconceived response to his need for attention from others; a response that may have been reinforced by the reactions of others. It is possible to imagine how a child, because of the way it is treated and the limitations on what it knows, comes to see spiteful behaviour as the most appropriate means to satisfy its need for attention. To imagine this is to see how we ourselves would respond were our horizons as limited as the child's. Once the spiteful characteristics are acquired they could become significant features shaping the way the child perceives and understands the world and guiding the way it responds. It is possible to imagine all the steps in a life history leading from an innocent child to a spiteful adult as ones we ourselves

would have taken had our vision been similarly limited. Spite and malice, therefore, can be understood as resulting from limitations of vision and from inferences drawn because of those limitations; and this, indeed, is the way in which such cognitive and imaginative limitations have been understood in the long tradition of thought about wrongdoing from Plato, through Spinoza, to Freud. It need not, therefore, be the idea of behaviour as law-governed, ***per se,*** that has led Spinozists to doubt the concept of moral responsibility, but the thought that bad character and bad behaviour can be understood as issuing from limitations of vision (themselves arising from heredity and environment) which would have led any of us to take just the same steps.[13]

It might be thought that if the real force of Spinozist objectivism is in the idea that wrongdoing issues from acquired limitations of perspective, rather than the notion that it is causally necessitated, determinism is no longer an issue.[14] This is true in a sense, but since the limitations we are considering are thought to be ***caused*** by heredity and environment, and therefore have a certain relation to the idea of determinism, the difference it makes is not such as to require a major revision of the standard ***dramatis personae*** (compatibilists, incompatibilists, libertarians, etc.) in discussing moral responsibility.

At any rate, the argument about limitations of perspective is a difficult line of reasoning for compatibilists to resist, given their acceptance that ignorance, illusion, and so forth all provide reasons for withdrawing or tempering the reactive attitudes and withholding blame. For example, someone who behaves in a malicious way towards me may be excused if his malicious attitude was based on a mistaken, though understandable, view of my actions. In that case his malice would have been directed at non-existent motives and, in a sense, at a non-existent person. The mistaken perspective of someone who is characteristically malicious may be so deeply ingrained that it is difficult to isolate or dislodge, and such that their whole personality may be structured around it; and there is no obvious reason why it should not be viewed as removing responsibility for their actions.

None of this exactly undermines Strawson's arguments about the reactive attitudes. The fact that a person's limitations of perspective can count against blaming them for what they have done, and that all bad behaviour might be attributable to limitations of perspective, does not entail that no one ever knowingly does wrong (even if Plato, Spinoza and Freud may have thought that it does). Someone's acquired limitations of perspective may incline him to think and feel in characteristically despicable ways, but he may yet be a knowing sinner. The link between acquired limitations of

perspective and bad behaviour is no reason for thinking that all bad behaviour must be excused, but it does leave us with doubts in many cases about how much weight we should attach to it, and these doubts can lead to a slippery slope giving rise to feelings of insecurity about blaming in any particular case. If we accept that all bad actions result from limitations of perspective and if we consider limitations of perspective as excusing wrongdoing in some cases, then why not in others? and why not in all?

JUSTIFICATION

We face an impasse. The practice of holding people morally responsible and the family of reactive practices and attitudes of which it is a part are, seemingly, ineradicable from our natures and our dealings with one another. Taken as a whole they cannot be justified, but neither do they require justification. Nevertheless, in any particular case of wrongdoing we face a choice about what attitude to take. To take the objective attitude is to forswear blaming and praising as we ordinarily understand them. And yet there are reasons for adopting the objective attitude that Strawson does not consider, and no compelling reason for refusing to adopt it. Our feelings are very much at issue in all of this, but our feelings are not decisive: we must decide.

One way of responding to the impasse would be to decide emphatically always to refuse the objective attitude unless it is unambiguously called for. To do so, however, would be no better than retributive humbug: a contrived response which would be guilty of ignoring human feelings in much the same way as those philosophers, criticised by Strawson, who ignore the human feelings that are expressed by our habits of praising and blaming.[15] The feelings that such a strategy would ignore overlap with those that inform the reactive attitudes; namely sympathy and a sense of fairness. It is our human feelings, therefore, that issue in the judgement that it is 'hideously unfair'[16] to blame someone for malicious behaviour that results from misconceptions and limitations of vision that would have led anyone to behave in the same way.

But if it is unfair, would we think it any less unfair if we thought that wrongdoers were possessed of libertarian free will? According to those who are sceptical about moral responsibility, even though they consider libertarian free will to be inane, the answer is yes, or, at any rate, that the concept of desert, which can be explicated as 'fair reward' (whether for good or bad behaviour), presupposes that of libertarian free will.[17] Presumably what has led them to this conclusion is the thought that it is impossible to

think of behaviour we regard as law-governed as also being blameworthy. Setting aside, for the moment, the argument that it is not the notion that all human behaviour is law-governed that is really behind scepticism about moral responsibility, is it true that the concept of desert presupposes that of libertarian free will? Consider the following propositions about Bob's act of deceit: 'he acted of his own free will', 'it was his doing, and he was wholly and solely responsible for it' and 'he categorically could have done otherwise than he did'. Suppose we stipulate, if there is any doubt about it, that these propositions are all to be construed in the sense that incompatibilists believe to be required for moral responsibility: does any of the three propositions, taken on its own, logically imply or in any way entail that Bob deserves to be punished or blamed for what he did? The answer is clearly that none of them do; and supposing that there is something unfair about blame and punishment, none of the propositions would entail or imply that they are any less unfair.

Is asking for a relationship of entailment or implication between propositions about responsible actions and desert demanding too much? Well, if not that, what else would establish the connection? Incompatibilists and sceptics might fall back on saying that it is a constitutional fact about us, or an entrenched conviction, that we are disposed to take desert as attaching to actions of the kind described in the above propositions. But then of course they will be open to the same questions they have pressed against the Strawsonian account, namely, 'are we ***right*** to believe that there is such a connection?' or 'is it ***fair*** that we make it?' If they claim instead that there is a conceptual dependency between desert and libertarian free will and that any explication of ***desert*** that makes no reference to libertarian free will is not an explication of what they mean by ***desert,*** we may ask what makes them so sure that that is what they mean, and how they know that it is what everyone else means or what is always meant. And supposing that they could find plausible answers to those questions, would they want to say that the semantic tie-up makes blaming and punishing any fairer than they would be without it?

What makes it seem as if the concept of desert requires libertarian free will is awareness that consideration of the etiology of behaviour undermines the tendency to blame (especially, as I have suggested, awareness of the thought that bad behaviour arises from limitations of perspective which themselves have formative and ongoing causes). To think of a person's actions as free in the libertarian sense is to think of them as lacking any etiology (or that some element in their initiation does). That may make it seem as if the thought that someone deserves to be punished, which was undermined by considerations about causes, is justified if their

behaviour has no causes. But the notion that someone deserves to be punished is no more justified if their behaviour lacks a causal history than if it has one. The conviction that punishment is fair reward for bad behaviour - that the guilty should suffer - is foundational to our moral thought and practices.[18] And since any moral justification must be internal to our moral thought, there is nothing that could provide a moral justification for that belief.

If desert, punishment and treating individuals as morally responsible for their actions cannot, as practices, be given a justification, what is it about them that might be thought of as monstrously unfair? The unfairness, I suggested, is thought to attach to the idea of seeing malicious behaviour as the product of limited horizons and misconceptions that would have led anyone else to behave wickedly. Presumably it would not be thought of as quite ***monstrously*** unfair if instead of people being punished for bad behaviour, good behaviour was always rewarded. What is thought to be monstrously unfair is the infliction of suffering on the wrongdoer who has done what we ourselves would have done given the same psychological circumstances. That we consider there to be something monstrous about the infliction of suffering reflects another part of our moral thinking and practices. It is the idea or attitude that suffering, pain, disease, and so on are bad and that pleasure, happiness, well-being and so forth are good. This is the part of morality, in other words, that utilitarians wrongly take to be the whole of it. Let us call it ***the principle of well-being.***

Libertarians and retributivists have not of course felt obliged to provide a justification for punishment, desert and moral responsibility according to the principle of well-being. It is the principle of well-being - involving the thought that causing suffering is, ***ceteris paribus,*** a bad thing to do - nevertheless, which has led them to think that these practices and concepts stand in need of a justification; albeit one that would place the practices beyond the domain of the principle of well-being.

My account of what has led philosophers to suppose that the concepts of desert and moral responsibility presuppose that of libertarian free will can be summarised as follows. We face choices in respect of particular actions, about whether to blame and punish or to understand, or, as we might say, about whether to adopt the reactive attitudes or the objective attitude. The idea that any bad action results from limitations of perspective gives us a reason (albeit, not a conclusive one) for thinking that the objective attitude is the appropriate one to adopt, since it seems inconsistent to blame or punish somebody for doing something that, given the same limitations of perspective, is what anyone else would have done. The fact that blaming and punishing inflict suffering, which conflicts with the principle

of well-being, leads us to seek a justification for blaming and punishing. We notice that the impulse to blame and to punish is undermined by thinking about the etiology of behaviour. What undermines the impulse is that conceiving of the etiology of behaviour necessarily involves adopting the objective attitude towards it and that to adopt the objective attitude is, for as long as one does adopt it, to relinquish the reactive attitudes. It seems, nevertheless, as if the loss of the impulse to blame and punish, when we think of wrongdoing as having a causal history, comes from a judgement that blaming and punishing are unjustified (rather than just incompatible with taking the objective attitude). It is inferred from this that what renders the practices unjustifiable is the thought that all actions have sufficient antecedent causes that render them explicable, intelligible and predictable. This in turn appears to entail that the applicability of blame and punishment, and of related concepts such as desert and moral responsibility, presuppose that it is false that all behaviour has sufficient antecedent causes - that some behaviour is the product of libertarian free will - and that if this presupposition can be satisfied, the threatened practices and concepts can be justified. But libertarian free will does not justify either these practices or the concept of desert, since it does not entail or imply them and since, in any case, their unjustifiability did not follow from determinism but from the fact that their place in our moral thought is foundational and therefore that they can neither be given a justification nor stand in need of one. A justification is therefore sought for something that does not require and cannot be given a justification.

The impulse to find a justification for punishment, desert and moral responsibility, therefore, comes from judging one strain or principle in our moral thinking and practice by another. If we put the matter like that, we may now ask whether the principle that the guilty should suffer needs to be justified just because it conflicts with the principle of well-being. Both are foundational to morality[19] (as we know it) and there is no reason to accept that one should be ranked lower than the other as a moral principle, still less that one should have to be justified because it is thought to conflict with the other. The thought that punishment, desert and moral responsibility require a justification is therefore a type of category mistake. It involves the employment of the principle of well-being beyond its proper sphere of application. The principle of well-being does not extend to the guilty, or, rather, it is qualified in respect of them. Neither, it may be added, does it apply, without qualification, to enemies in war. Nor can it provide an *a priori* justification for redistributing private property to those who will benefit most from it.

Belief in the universal applicability of the principle of well-being is one

of the dominating illusions of our age. It involves turning a blind eye to morality's darker side; the part of morality that views the suffering of the wicked as good or right (regardless of good consequences) and in which the application of the principle of well-being is restricted. In averting our eyes from it, we tum away with a shudder from a central feature of our moral selves. But our understanding of morality will always be skewed and unreal for as long as we fail to face up not merely to our own desires for the suffering of the guilty, but to our fundamental conviction that it is right that they should suffer.

THE IMPULSE TO JUSTIFY

The foregoing account of the place of punishment, desert and moral responsibility in our moral thought and practices is, I think, important and true. The belief that desert and punishment, as such, require a justification has no obvious validity and there is no reason, therefore, for us to regard the practices of blaming and punishing, as such, as monstrously unjust. But is the argument persuasive enough for us to continue confidently to blame and punish, just as long as the usual criteria for the applicability of our reactive attitudes are satisfied? There are two related reasons for thinking that it is not.

In the first place, as I suggested above, the argument that our moral practices and judgements are bound up with attitudes and feelings, rather than metaphysical judgements, cuts both ways. The principle of well-being is no more justified in the light of some prior judgement that pain and suffering are self-evidently bad, than the principle of desert would be justified by our possessing libertarian free will. Our judgments that pain and suffering are bad, as Hume rightly observed, are grounded in our feelings of sympathy: the capacity to be moved by the fortunes and misfortunes of others. The woman's judgement that her partner's spiteful words were brought on by stress and overwork does not simply repress or qualify her reactive feelings towards him, but expresses her feelings of sympathy and concern with his well-being. The Spinozist's judgement that it is inappropriate to hold the wrongdoer to blame for behaviour that was understandable and predictable, given his cognitive circumstances, is likewise informed by his or her concerns about suffering humanity.

The fact that it is feelings or moral sentiments that underpin the principle of well-being, just as much as the principle of desert, undermines the claim that the impulse to find a justification for punishment and desert is a straightforward category mistake. To say that the scope of application of

the principle of well-being is limited by other principles, such as desert, private property, obligation to one's own community and so forth, is to imply that there is some rule or understanding that stipulates this. But there is no such rule and no such understanding. There is, I agree, a category mistake (that of making one foundational principle subordinate to another) in arguing that desert and punishment require a justification. But this is not to say that there are no grounds for being troubled by the suffering caused by punishment and blame, and it begs the question in favour of the principle of desert to suggest that it delimits a domain in which the principle of well-being has no place. The truth, as we know from the case of the woman who tries to be understanding about her partner's spiteful remarks, is that in many cases it is possible for us, rightly, to blame and punish, but also to decide instead to set aside the impulse to do so and to try to understand and accommodate.

Likewise, while it is true that utilitarian claims that we should be no less concerned with the welfare of our enemies than with that of our friends are a distortion of our moral thought and practices, we do nevertheless have moral concerns about the treatment of enemy soldiers and citizens, which express our feelings of sympathy with other human beings. It is possible to repress those feelings, and to argue that certain situations positively require us to do so, but it cannot be maintained, at least not by those like Strawson who emphasise the connection between moral judgements and practices and our natural human feelings, that moral concerns about the treatment of enemies are just confused. And again, while it is mistaken to think that respect for private property must be justified according to the principle of well-being, there is no mistake in thinking that respect for private property should give way to the principle of well-being in particular cases (even cases where large-scale nationalisation is envisaged), or even in thinking that circumstances might develop in which the cumulative effect of decisions taken about what to do in particular cases brings about a sea change in our attitudes about private property.

One argument that might be pressed against this is that Strawson's account of the reactive attitudes and feelings situates them within personal and other relationships and makes them dependent on our expectations of goodwill and consideration within those relationships. Our natural human feelings do not constitute an unquestionable bottom line for answering questions about our moral and interpersonal practices.[20] These feelings have their proper place and expression within certain relationships. It may therefore be argued that a concern for the suffering of others, simply on account of a shared humanity, isolates our moral sentiments from their meaningful situation within social relationships. I agree that interpersonal

and wider relationships are an enormously important constitutive feature of morality, and perhaps a convincing argument could be made that such relationships place constraints on the meaningful scope and application of the moral sentiments. Nevertheless, the suggestion that we have no moral connections, and no proper objects of moral sentiments, other than to those with whom we have definable social or personal relationships is quite implausible. Were such a view generally to be taken to heart, it would constitute a major revision in our moral thinking. Most people consider that there are moral constraints on how enemies should be treated, to say nothing of our obligations to assist distant peoples in circumstances of famine, ruin and poverty.

Thus, the impasse I identified above, about whether in any particular case to respond objectively or reactively to wrongdoing, reflects deeper tensions in our moral thought between the principle of well-being and other foundational principles, especially the principle of desert. These tensions, moreover, are grounded in conflicting feelings or sentiments we all (or almost all) share. Whether to respond reactively or objectively to wrongdoing - whether to blame or to try to understand - is a question for which there is no general answer. Our own moral thought, and our own natures, pull us in opposing directions. There is no reason universally to abandon our practices of praising and blaming, even supposing that we could; but there is no absolutely compelling reason in any particular case for holding anyone responsible or engaging in blame.

The problem of whether moral responsibility is compatible with determinism, therefore, is misconceived. The philosophical inclination to ask whether moral responsibility is compatible with determinism assumes that it is possible to justify the practice of holding individuals responsible for their actions, but no general justification is either possible or required. The problem is rather that we face a decision in respect of any particular instance of wrongdoing about whether to blame or to adopt what Strawson has described as the objective attitude.

When we engage in the objective attitude and consider the causes of bad behaviour, the idea that such behaviour arises from misconceptions and limitations of perspective that would have led any of us to do the same, and the fact that it is impossible in the face of such a thought to engage in genuine blame or reactive feelings, can lead us to suppose that if this is true the reactive attitudes are based on falsehood and illusion. To take that view, however, is to beg the question in favour of the objective attitude. The truth is that it can be open to us to adopt either attitude and that it is not within our power to completely relinquish either attitude. We stand on uncertain ground. With no clear guidelines about what attitude to

adopt in any particular case, we are confronted with a persistent moral tension between the impulse to blame and the desire to understand.

It is arguable that the tension between the principle of well-being and other features of our moral thought and practices has been sharpened by social changes such as the emergence of the welfare state, the growth of social science and so on. It is conceivable that if the principle of well-being assumes greater importance for us, moral responsibility, desert and punishment will no longer assume the same proportions for us that they have hitherto (although, as Strawson has argued, an overnight shift away from reactive attitudes and practices is barely conceivable). At any rate, the tension we face about whether to blame or to understand is not, as such, philosophical in nature or open to philosophical resolution. The tension is an inescapable feature of moral life, and if it ever can be resolved, it will be as a result of social and moral evolution.

The real force of the problem of moral responsibility is not about what is meant by ***morally responsible*** or about whether holding people accountable, and blaming and punishing them, are in general justified; and neither does it stem from contrary metaphysical attitudes about how blaming, resentment and desert stand in relation to determinism. There is, as such, no problem about the relationship between moral responsibility and determinism. The true problem is a practical one about opposing strains within our moral framework and conflicting (non-metaphysical) moral sentiments within ourselves. It is a problem for which, as things stand, there is no general solution.

Despite the lack of a general solution or answer to the problem, I think that there are considerations that can guide the individual in deciding whether to blame or to understand (assuming that blaming and understanding are always exclusive). It would be wrong, of course, to think that whether to blame or understand is always something we make decisions about, but we sometimes do: we are not always simply carried along by our feelings in these matters. It is our feelings, however, that do, on the whole, guide us in making such decisions, and the feelings in question have to do with our interests (what we care about) in the behaviour of others. The reactive attitudes, according to Strawson, are structured by our expectations of and desires for goodwill and consideration on the parts of others with whom we are involved in interpersonal relationships of various sorts. I have argued that the objective attitude is also often connected with our interests in or feelings about the fortunes and sufferings of others.[21] The objectivity of a therapist or doctor is informed by a primary interest in a client or patient's well-being, as opposed to desires for their regard.[22] The woman who resists the impulse to respond with resentment and blame

to her partner's spite has both kinds of interest in his behaviour. She values his goodwill and consideration while also being concerned with his well-being. Clearly these interests are not opposed but interdependent and mutually sustaining. But when the person in whom she has these interests behaves badly towards her, they prompt conflicting responses. Her attempt to understand her partner's bad behaviour reflects not only an interest in his well-being, but also a judgement that her interests in his goodwill and regard are not seriously challenged by uncharacteristic outbursts.

One can be led in the direction of the objective attitude, or to modify one's reactive attitudes, by a range of considerations and interests. The interest one has in the regard and goodwill of others might itself lead one to judge that the objective attitude is appropriate, where protest and remonstration have persistently failed to alter another's hurtful behaviour. In that case we might come to view the other person as incapable of responding to our hurt feelings. We would judge, in such cases, that there is a cause of the persistent behaviour, and in doing so we would be drawn into the objective attitude.

There are considerations that for some of us would provide reasons for responding objectively, while leaving others unmoved. Someone who is committed to the view that poverty and inequality are evil has an interest in the misfortunes and suffering of those who are disadvantaged, and this will predispose her towards the objective attitude in respect of their wrongdoings and the causes thereof. A liberal free marketeer, we may expect, would not be thus predisposed. It is possible that there are individuals - Spinoza may have been one - who are so saintly and detached from interpersonal relationships that their desires for the goodwill and consideration of others are very attenuated and their interests in the general well-being of humanity are unusually strong. Given what they care about, such persons will naturally tend to view wrongdoing objectively and without reproach. For them there will always be an interest in perceiving the sufficient causes of wrongdoing, and the thought that all behaviour has sufficient causes will be a constant background to their desires, pursuits and dealings with others. But there is no reason why one should want or seek to be like Spinoza, and there is no reason why one should always strive to remind oneself that all behaviour has sufficient causes.

OURSELVES AND THOSE CLOSEST TO US

When it comes to one's own misdeeds the argument that decisions about whether to blame or understand are informed by one's interests is rather

less straightforward. Inclinations to feel guilt, regret and self-recrimination involve feelings not only of concern for those who have been hurt by our actions but also that in acting as we did we have demeaned or devalued ourselves. The interests that can lead us to view our misdeeds objectively include feelings of being overburdened by self-recrimination and the desire to avoid such behaviour in future by understanding what caused it. For some people the desire to be free from self-recrimination can be so dominating as to result in a persistent refusal to take responsibility for their actions. Others, by contrast, find it especially difficult to view their misdeeds objectively and only ever partially and temporarily succeed in doing so. I take it that this inability to view one's own wrongdoings objectively is more common than the trait of always ducking responsibility and making excuses. It is not entirely clear to me why this should be so. Possibly the inability is motivated by pride and distaste for having demeaned oneself rather than concern with others. Pride in oneself naturally requires a sense of control over one's actions and would not be well served by persistent refusal to take responsibility for one's shortcomings (unless one has some means of avoiding awareness of the contradiction). Another suggestion is that it is not possible to take an objective view of one's current and future actions and their causes, and that there is a consequent dissonance involved in adopting the objective attitude towards one's actions when they are past. Either of these explanations could be true, but I see no obvious reason for accepting them. I remain unsure, therefore, why many of us are more deeply attached to the reactive attitudes in respect of our own actions than those of others. All that this difference need lead us to suppose, however, is that our interests and feelings about our own actions are rather more complex than in respect of the actions of others. There is no reason to think that decisions about whether to blame ourselves are any less informed by our feelings and what we care about than those about whether to blame others.

Our feelings and interests, as Hume observed, are naturally stronger in respect of what lies closest to home.[23] If they are our guide in whether to blame or understand, we will be more inclined to blame where our personal interests and those close to us are threatened. Only to hold responsible those who have harmed us, or those close to us, would be unjust and morally incoherent. This might seem to support the view that to hold someone morally responsible is properly grounded in a metaphysical view of them and what they are capable of, rather than in our feelings and interests in their behaviour. In fact, of course, people do tend to be partial and inconsistent in their practices of praising and blaming, but the inconsistency in their doing so does not undermine the account I have given.

The account of whether to blame or understand as following from one's interests or what one cares about does not imply that it is solely what one cares about, in each case, that informs whether to praise and blame, but rather that it determines the principles according to which we praise or blame. To blame someone for her misdeeds is to adopt a view of those misdeeds: a view that is prompted and structured by one's interest in being treated with proper regard or for others to be thus treated. To blame with only the injury to oneself or another in mind, and without regard to any standard or principle of proper treatment, would be not to have a proper view of the object of blame at all. The objectivity of an egalitarian who forswears indignation at acts of vandalism and burglary until they are visited on her will rightly be regarded as more reflective of indifference to the suffering and loss of other victims than concern about the disadvantages of the perpetrators.[24] A therapist's objectivity is not conditional on the degree to which any of her clients have caused harm or disadvantage to herself, her family or her friends. If she finds herself unable to resist the inclination to respond reactively to a client's behaviour, the proper course of action is to discontinue the professional relationship. That decisions about whether to blame or understand are prompted by our feelings and interests does not entail that such decisions are inevitably partial and inconsistent.

CONCLUSION

I conclude, therefore, that the traditional problem of whether moral responsibility is compatible with determinism is misconceived. The real problem is a practical one about whether, in particular cases, to praise and blame or to understand: a problem that does not arise from entrenched and conflicting metaphysical attitudes, as Honderich and others have supposed, but from tensions in our moral thought and practices.[25] There can be no general resolution of the tension between the principle of well-being and the principle of desert, and therefore we as individuals are inescapably confronted with conflicts of intuition about whether to blame or to try to understand particular instances of wrongdoing. Nevertheless, if we keep in mind that it is what we care about, informed by our personal, moral and political feelings and sentiments, that generally informs whether we take the objective or the reactive attitudes, we will not be faced with a hopeless dilemma every time we confront decisions about whether to blame or to understand.

Having come to an understanding of the problem of moral responsibility

that resolves (or dissolves) the issue of its compatibility with determinism, I turn now to those problems of free will and agency that, as I argued in the previous chapter, can be distinguished from the problem of moral responsibility.

3 Free Will

There are a variety of conditions and circumstances that may lead us to judge that an action has not been freely performed or that a person is not free. Very often the circumstances have to do with being obstructed or frustrated in doing what one wants or in attempting to bring something about. Imprisonment, censorship, lack of access to political office or participation are all thought to limit freedom, as are lack of funds, denial of career opportunities, and social isolation. There are, in addition, various 'internal' circumstances, including neuroses, compulsive behaviour patterns and weakness of will, as well as limitations of skill and intellect, that can thwart our hopes and desires. What lack of freedom comes to in respect of these and similar circumstances is not being able to do or bring about what one wants.

In Chapter 1 I noted that the idea of freedom as being able to do what one wants appears to be at odds with a familiar idea or image of free will having to do with the experience of being able to choose or decide for ourselves, especially where practical reasoning offers no clear guidance about what to do: the liberty to choose even what we do not want. The perennial clash between compatibilism and incompatibilism appears to owe a lot to these different ideas of freedom, even if they are not its only source. The seeming incommensurability of the two ideas lends support to the claim that we have opposing or irreconcilable attitudes about free will and the initiation of actions, with respect to which the meanings and uses of the key terms are ambiguous.

In order to assess the claim that our feelings and attitudes about free will are inconsistent or irreconcilable, I propose to begin by discussing what is thought to be required for the first kind of freedom - 'the liberty of spontaneity'[1] - and then to consider whether it is consistent with what we would otherwise say or feel about free will.

FREE WILL AS DOING WHAT YOU REALLY WANT, BECAUSE IT IS WHAT YOU REALLY WANT

As I remarked in Chapter 1, few nowadays are very much persuaded by Hobbes's classical compatibilist account of free will as the ability to act on one's strongest desire. A thoroughly Hobbesian human nature would leave little room for any independence of mind or action in respect of

strong desires and could not accommodate the idea that we can be unfree in being unable to resist unwanted desires. Acting on one's strongest occurrent desire is often enough not acting as one would want.[2] We might say, then, that free will consists in being able to do what we ***really want.*** But if what we really want can be different from what we most desire, how is it different? Contemporary 'hierarchical compatibilists' have sought to answer this question by specifying a motivational structure for free will, according to which free agents are those who have a capacity to choose or identify with some desires over others with which they conflict and to act according to such choices or identifications. Suggestions about how we might be said to identify with certain desires include our having higher-order desires in respect of them or making evaluative judgements about them, but these details need not hinder us just now.[3] If a person acts to satisfy a desire with which she does not identify, she does not do what she really wants to do: she acts unfreely.

It is possible, however, for a person to do what she (really) wants, and yet fail to do so freely. It sometimes happens that what we want to do coincides with what we are compelled or required to do. An absent father who wants to provide financial support for his children may be legally required to do so, so that he would have to make the payments whether he wanted to or not,4 and we would not naturally describe his actions as free. Let us say then that free will consists in being able to do what one really wants ***because***[5] it is what one really wants.[6]

There are many situations in which knowing what one really wants is reasonably straightforward, but often enough it can call for work and a degree of skill in being able to reason practically not only about one's goals and how to get them but about what one really cares about and what kind of person one wishes to be. This in turn calls for qualities such as intellectual organisation, determination and critical detachment.[7] Determination can also be required in order to put one's judgements about what one really wants into action; and also the ability to develop habits of thought and action that will minimise the effects of recalcitrant desires and weakness of will. All of this has long been recognised by a philosophical tradition that sees freedom not as something we naturally possess in being able to act otherwise than we do, nor as a feature of favourable circumstances, but as something we ***acquire*** by developing right and reasonable habits of mind and behaviour, including the ability to view our fortunes with calm and detachment. The tradition is typically associated with the Stoics, although it predated them and has long outlived them.[8]

Having free will, it would seem, is no straightforward matter. Being able to do what one really wants because it is what one really wants

calls for proper motivation, rationality, detachment, determination and generally good habits of thought and behaviour. But how much and in what degrees are these qualities called for? If we set the standards too high for what is to count as free will, so that it can be enjoyed only by the wise and venerable, doesn't this conflict, as the Stoic conception has often been thought to do, with what we would otherwise be inclined to say, which is that most of us have choices in what we do and can be more or less free in making them?[9] And if there is a conflict, shouldn't we then accept that our prereflective attitudes about free will are irreconcilable or incoherent? and that free will is a confused or illusory ideal?

The answers to these questions will depend, as I suggested in Chapter 1, on whether, in the first place, a closer examination of our apparently conflicting attitudes and intuitions can show them to be reconcilable and whether, in the second place, it shows them to be veridical or illusory. The answers in respect of moral responsibility were that there is no irreconcilable conflict within the reactive attitudes of which it is a part and that it does not presuppose metaphysical images or propositions that would be false should determinism be true. My examination of whether there are conflicting attitudes or intuitions about free will, other than in respect of moral responsibility, begins with the suggestion that acting freely involves doing what is morally right.

MORALITY AND FREE WILL

The convergence of the Stoic tradition and contemporary compatibilist accounts of free will is noteworthy, given the claim that our prereflective attitudes about free will are divergent and possibly incoherent. The convergence can be overstressed, however: most compatibilists would be resistant to any identification of free will with doing what is morally right, especially if this is linked to any suggestion of moral realism or objectivism. But it is also important that apparent differences should not be exaggerated. Members of the Stoic tradition have been careful to emphasise that a free person who acts in obedience to what is right does so because of reasoned conviction rather than fear.[10] They have also argued that to do the right thing because it is the right thing is to realise one's true or essential self, in opposition to enslaving desires and habits. Again, this parallels contemporary compatibilist accounts of free will as the ability to act on those wants with which we identify.

Still, if the Stoic conception of freedom commits its supporters to moral realism, then it carries a heavy burden. If free will consists in doing the

right thing - not just what is reasonable - would free will then be an impossibility in the absence of objective virtues and vices? Or, instead, would disagreement about what is right entail disagreement about whether a person or an action is free? Even supposing that moral realism might turn out to be true, a resolution of the problem of free will would be very unpromising if it awaited a conclusion to the argument between moral realism and its opponents.

If freedom of will is linked to doing what is morally right, realistically understood, this excludes the possibility that evil actions or evil people can be free, and this, as I mentioned in Chapter 1, has been the view of many members of the Stoic tradition. The idea was contemptuously dismissed by Bentham:

> Is not the liberty to do evil liberty? If not, what is it? ... Do we not say that it is necessary to take away liberty from idiots and bad men, because they abuse it?[11]

So what has led philosophers to believe that the wicked cannot be free? Susan Wolf, who is the most well known contemporary advocate of this view, asks us to imagine a hypothetical dictator, JoJo, who inherits power from his evil father and who, having been brought up to be cruel and lacking in the ability to distinguish between right and wrong, cannot be held responsible for his acts of cruelty. At the same time, Wolf argues, he is capable of valuing the way he is and being able to continue to act as he does. He really wants to be the way he is and therefore satisfies our criteria for free will. The capacity for reflective evaluation and the ability to determine one's actions in accordance with it are insufficient, Wolf thinks, for moral responsibility.[12]

Wolf's argument is clearly related to the view I discussed in the previous chapter, that people should not be held responsible for bad behaviour that issues from limitations of perspective; and it would be perverse to deny that someone who cannot grasp the difference between right and wrong cannot properly be held morally responsible for what he does.[13] It does not follow from this, however, that such an individual could not have free will.

Most of us have need of the good opinions and consideration of others and are capable of feeling deeply moved by what happens to others. It is inescapable, therefore, that since 'our values' refers to the things we care about, those needs and feelings will necessarily inform our values. If having free will involves being able to judge our actions and desires against our values and in being able to put those judgements into action, it must also be true that actions that directly conflict with our needs and feelings will

also conflict with our values and with any judgements on which they are based. Therefore, behaviour like JoJo's (behaviour that issues from relentless cruelty), for creatures such as ourselves who are interested and concerned about others, could not be free.

Is JoJo like us? Clearly he would be capable of acts that most of us could never commit. But would he care how others feel towards him? Would he have need of recognition, respect and love? If he would, then his needs and cares would conflict with his cruelty and manipulativeness in a way that would make it necessary for him to determine what matters most to him and to root out those conflicting feelings and behaviour that are less important to him. If he could not do that, he could never make unambiguous evaluative judgements about what to do and therefore could never be free in acting on any of his conflicting values and desires (and the same will hold for any of us, where we are unable to resolve deep conflicts of wants and values). But suppose he could, and that he concludes that his needs for love are less essential than his ingrained cruelty; and suppose he were to successfully transform himself in line with this judgement; or suppose that he never has such needs: in that case nothing need prevent him from determining what matters most to him or from evaluating his desires and behaviour in accordance with it. By our standards - our human standards - he would be a monster, but I can see no reason for denying that he could be free.

It is worth adding that there is something rather odd about asking whether a creature such as JoJo is one whom we would hold responsible for his actions, and also about the notion that cases like JoJo's might have any direct bearing on our practices of blaming and holding people accountable for their actions. What we would naturally want to do with someone like JoJo is to depose him and place him under lock and key as quickly as possible without troubling ourselves about whether he is truly responsible for what he does.[14]

Many of us might be inclined to think that there never could be such a creature; and that the truly infamous have, without exception, been deeply at war with themselves; having human needs they were never able to fulfil. But the possibility of a creature who lacks such needs and can, with a free will, commit acts of wickedness cannot be ruled out *a priori.* If there are constraints on what we can judge to be valuable or worthwhile, it is because our humanity places constraints on what we can care about and not because it is 'true' that virginity is a virtue or that it is 'a fact' that sons should not desert their ailing mothers (or whatever else).

To say this is not exactly to reject moral realism. If there are objective moral values and they are true regardless of what anyone says or does,

then to the extent that free will consists in doing the right thing, it does so because our human needs stand in some relation (whether one of dependence or of happy coincidence) to those values. Either moral values are necessarily bound up with human needs, in which case there is a direct relationship between free will and doing the right thing, or they are self-subsisting truths, in which case there is only an indirect relationship between free will and doing the right thing.

What follows from this? Wolfs argument that JoJo could not be free is directed against contemporary compatibilist claims that to be free is a matter of being able to govern one's behaviour in accordance with one's ***real self*** or, in other words, with what one really wants. Wolf's view is that one must also be able to judge one's real self and to bring it into line with 'the true and the good'.[15] She rightly makes the point that it is not possible for just any old judgements or choices to be the basis of free will. We are human beings with human needs, and this places constraints on what it is possible for us to value and what it is possible for us to really want. But to suggest as Wolf does that free will consists in being able to revise our deepest values and wants in line with ***the true*** and ***the good*** suggests that truth and goodness constitute some sort of independent standard that the individual and her values must live up to. To revise one's values in light of what is right and true, however, does not require going outside one's evaluational system to some external standard. ***The truth*** that an agent must measure her wants and values against is what the world will allow, including her own human needs and feelings. And inasmuch as she judges her values against ***the good,*** the good exists for her as part of her values. The kind of free will we can enjoy is bound up with our human natures. To repeat: the argument about moral realism is a red-herring. Our human needs place constraints on what we can consider to be good or worthwhile. If, as moral realists suppose, these judgements might conflict with what is somehow objectively good, then what is objectively good would stand in no direct relation to free will.

FREEDOM OF ACTION AND FREE WILL

There are situations of which we might be inclined to say that a person is free in one important sense but unfree in another. A person may be said to be 'compelled' at gunpoint to do something he would rather not, but also to have 'freely chosen' to comply, in a sense that is not thought to be true of addictive or neurotic behaviour. It has been suggested that our ordinary grasp of the distinction between these two senses of freedom is

rough-and-ready and inconsistent, and that any attempt to make it consistent and precise, while theoretically useful, will be necessarily arbitrary.[16] This, if true, would give further weight to claims that there are fundamental inconsistencies in our prereflective beliefs and intuitions about freedom.

According to Harry Frankfurt, a bank clerk who agrees to the demands of an armed robber will do so freely, even though he may desire not to do so, if be also has a desire to save himself and a *second-order volition* (a desire that some first-order desire be effective) that it should be this desire that moves him to action.[17] If instead the clerk acts as he does, not because of his second-order volition, but because his fear is so strong that be finds compliance irresistible, his action will not be free. Frankfurt concedes that to say that 'some actions may be petformed freely even when they are petformed under duress' has a jarring sound,[18] and also that an adequate theory of free action would allow that someone who has acted under duress can be judged not to have acted freely. Frankfurt relies on the distinction between *free action* and *free will* to resolve the problem, claiming that one may act with a free will while acting under duress, even though one's action is not free. He concludes that there are simply two conflicting requirements for actions to be described as free, either of which might inform a univocal philosophical usage of *free action.*[19]

Clearly, however, any decision about what is to count as a univocal philosophical usage of *free action,* in addition to being arbitrary, would not address the problem of why we appear to have conflicting intuitions about whether to describe actions petformed under duress as free (in addition to which we might ask what the point would be of any decision about a univocal philosophical usage, and who would make it). A further reason for rejecting Frankfurt's solution is that as things stand, if we allow that someone who acts under duress can be unfree because of his circumstances, we will face problems in saying why highly attractive offers, including bribes, are not thought of as rendering us unfree.[20]

I think it can be shown that the asymmetry between the perceived effects of bribes and threats can, as Frankfurt suggests, be understood in terms of freedom of action as a distinct category from that of free will, without inviting objections that positing this second category of freedom is a merely arbitrary or ad hoe move. Freedom consists in being able to do what one wants because it is what one wants,[21] and a person's ability to do what she wants can be restricted by either external or internal circumstances.[22] The two categories of freedom relate to these two sets of circumstances. A person may therefore lack freedom of action in respect of what she wants to do, because her external circumstances prevent it, or

she may lack a free will to do as she really wants, as a consequence of an addiction or neurosis.

Judgements in respect of both categories of freedom, moreover, have a necessarily normative dimension, founded on what is reasonable, without which claims about freedom of action would be unintelligible. A country that allows freedom of action to representatives of the International Atomic Energy Agency's inspectorate, for example, will be understood to have granted rights of access, inspection, measurement, and so forth; or in other words, all those things the nuclear inspectorate could reasonably want, given its function. The country will not have granted permission for nuclear inspectors to blow up reactors or to organise fireworks displays for the workers.

There are infinitely many things a person might want to do, which the laws of nature, the limitations of human knowledge, lack of physical prowess, and spatiotemporal location would not permit. In consequence a person could claim that since all the things she would really like to do are denied to her by, say, her lack of supernatural powers, she never does anything willingly. Frankfurt tries to rule this out by stipulating that it is not enough for describing a person's behaviour as unwilling that she would prefer to be in a different situation, but that she must regret or resent her actual situation.[23] But suppose she does resent her mortal limitations, ought we then to regard her as unfree? Frankfurt suggests that a person's actions are unfree if what she does thwarts her higher-order volitions, or because the desires on which she acts are irresistible. If our frustrated mortal, knowing her limitations, has effective second-order volitions that she only acts on those desires that can naturally be satisfied, on Frankfurt's account we must judge her as acting freely. But there is a prior and commonplace judgement about her freedom of action which is indifferent to her motives, and which Frankfurt's account fails to articulate. In a trivial sense, of course, none of us are ***free*** to perform miracles, but this is not a significant restriction on anyone's freedom of action. We would not be prepared to accept lack of supernatural ability as a genuine restriction on freedom of action because our judgements about freedom of action are informed by what is humanly possible.

In addition to the limits of human powers, judgements about individual freedom of action also embody a sense of the average person's circumstances and needs. The average adult individual has a degree of movement that is denied to a prisoner, and we therefore judge imprisonment to be a significant restraint on freedom of action. The average person has various needs we would not ***reasonably*** expect her to ignore, and therefore if someone is offered a choice between having those needs denied and doing

something we might otherwise consider wrong or unpalatable, we do not think of her choice as free because choosing to do otherwise than she does, given her needs, is not something we could reasonably expect her to do. Thus the bank clerk's need to survive renders his choice of acceding to the gunman's demands unfree. It may be free in the second sense of being what he really wants to do, given his options, but it is significantly unfree from the standpoint of freedom of action.

The trouble with basing judgements about freedom of action on needs a person could not reasonably be expected to ignore is that it is famously impossible to arrive at any uncontroversial distinction between genuine and artificial needs.[24] It does not follow from this, however, that judgements about freedom of action are either arbitrary or merely conventional. There is disagreement about whether a person is 'free' in being able to choose to get a job or to spend her time in idleness. Some argue that such a choice is not free because a person has needs (not exclusively financial) she could not reasonably be expected to satisfy without a job. Those who deny this usually have some conception of absolute human needs, which are met for those out of work by benefit payments or charity. Identifying a distinct category or sense of free action from that of free will, however, does not require that we be able to unarguably resolve any conceivable disagreement about whether someone has acted freely, but merely that our judgements about free action are guided by intelligible and coherent normative criteria. That such criteria exist is shown by the many cases in which our judgements about freedom of action are not thus divided and by the fact that where there are disagreements they do not arise from the criteria themselves but from different understandings of the concepts they employ.

Therefore, one who accepts a bribe simply to satisfy desires for luxury and convenience does so freely, because her desires for luxury and convenience are not needs it would be unreasonable to expect her to deny (although if she suffers from some internal compulsion - say an addiction to luxury - that makes it impossible for her to resist taking bribes, we could not reasonably expect her to resist them). But a bribe or an offer that is taken in order to be able to eat, or to feed one's family, is not freely chosen, because the need to eat is not something one can reasonably be expected to deny.

Don Locke has argued that what leads us to say that someone is unable to turn an offer down cannot be that it would be unreasonable to expect them to refuse it, since if someone were offered £10 for a piece of junk, it would be unreasonable to expect them to turn it down but we would not call their choice to sell unfree.[25] Locke's example, however, does not compare like

with like. What would be unreasonable about expecting someone to turn down an offer that will save them from starving is not, as such, that given the options survival is preferable to refusing bribes, but that ignoring one's survival is something we could not reasonably expect anyone to do. The unreasonableness of expecting the person to choose differently derives from the unreasonableness of expecting her to ignore a fundamental need. The unreasonableness of expecting a person to refuse £10 for a piece of junk, however, is unrelated to the person's needs and is simply a matter of the utilities of the options in the choice. Certain choices are considered unfree not, as such, because it would be unreasonable to expect a person to refuse what is offered, but because it would be unreasonable to expect a person to refuse what is offered ***given her needs.***

We therefore have a concept of freedom of action relating to how things stand between what a person wants and what her external circumstances will allow her to get, which is distinct from the concept of free will (which relates to how things stand between what a person wants and what her motivational system will allow her to get). Any attempt to isolate a single sense in which an action can be judged to be free or unfree[26] that overlooks this distinction is therefore misconceived. It is important to know in what sense an action is unfree; but a fully free action must satisfy both senses. The apparent conflict in our prereflective intuitions about whether to describe an action as free can be resolved, therefore, without resorting to ad hoe philosophical decisions about which sense of ***free*** is to be taken as standard.

UNWANTED WANTS

If free will consists in being able to do what we really want because it is what we really want, it must follow that we can become free either by removing what stands in the way (externally or internally) of what we want or, alternatively, by ceasing to want it. The Stoics are notorious for believing that to be free one must accept and accommodate oneself to whatever is beyond one's power to change, even if that should include imprisonment or slavery. Thus, we might free a prisoner either by releasing him or by presenting him with a copy of Epictetus.[27] One who has acquired the habits of mind and action characteristic of a free man will be free whatever his circumstances. Therefore, according to Seneca:

> it is a mistake for anyone to believe that [the] condition of slavery penetrates into the whole being of man. The better part of him is exempt.

> Only the body is at the mercy and disposition of the master, but the mind is its own master and is so free and unshackled that not even this prison of the body ... can restrain it from using its own powers.[28]

A free person, according to the Stoic tradition, confronts frustrations and misfortunes with equanimity and detachment.

Commenting on this, Michael Slote suggests that 'the Stoic or Spinozist ideal of emotional detachment is an illusion for us, an ideal perhaps, but one that we are simply not capable of'.[29] **If** the bank clerk who is held up were to tell us that his attitude towards the hold up was one of calmness and emotional detachment, Slote argues, we would either have to conclude that he is incapable of loyalty or that he is refusing to confront his deeper feelings. Slote's criticism of the Stoic tradition is important, but as it stands it is open to challenge.

We are incapable of the emotional detachment that the Stoics claim to be possible, it seems, because we give love and have need of it, and giving and needing love means, in addition, that we must be capable of resentment, sorrow and anger. Not to have these feelings is to be incapable of love. Slote also points out, however, that the Stoic case for emotional detachment is based on the claim that many of our emotional responses are based on attachments to 'worldly things'. **If** 'worldly things' is taken to include other people and their goodwill, then it is hard to imagine that most of us would be capable of the detachment the Stoics advocate or to see why we should regard it as desirable. But if 'worldly things' is taken to refer to material goods, then the Stoics have a stronger case than Slote allows. **If** the bank clerk places a higher value on his life than money, he has a good reason to accede to the gunman's demands with equanimity. Slote's suggestion is that a second-order desire to behave heroically, and an attitude of loyalty, will lead him to be resentful and ambivalent at having to hand over the money. But if that were his reaction we might well regard him as being confused about what really matters in life, being guilty of vanity in wanting to play the hero's part and of misplaced loyalty in entertaining the idea that he should risk his own life for the sake of a bank's money. Even if we can imagine having just those feelings ourselves were we to be in the bank clerk's shoes, we would be wrong to regard them as either reasonable or inescapable.

To withdraw entirely from reactive attitudes and emotional responses, as I said in the previous chapter, would for most of us be unthinkable. It is not impossible critically to assess our emotional responses, and the beliefs, desires and values that prompt them, but it is often difficult and sometimes comes to nothing more than temporarily stepping back from

our initial feelings and asking whether they are really in order. At the other end of the scale from this is Spinoza's project of liberating oneself from the grip of ***passive emotions*** and ***inadequate ideas,***[30] and its counterparts in twentieth-century psychotherapy. How much you aspire to Spinozist detachment will depend on what kind of person you are and on what really matters to you. By all accounts Spinoza personified the detachment and calmness of mind that he commends, and in that case one is bound to regard him with admiration. If saintliness were ever a reality, then this, I suppose, is what it would be; and that is just the problem, for a saint is a little more (and a little less) than human.

But it does not require a Spinoza sometimes to be able to reflect and change how one responds to situations one cannot change, and even to modify what one wants. The idea that there is a chasm between Stoic/Spinozist detachment and what most of us are humanly capable of carries the tacit implication that there is a level of freedom denied to all but the wise and saintly few (even allowing that this level of freedom would, for most of us, be undesirable).[31] A better way of understanding the importance of reflective detachment is that the ability critically to evaluate the way we respond to certain situations is indispensable for free will, but that systematic and unvarying detachment and objectivity of attitude would undermine what makes reflective detachment worthwhile for most of us. If one's interests and attitudes are, as Spinoza's may have been, directed more at humanity and its condition than at those with whom one has more familiar relationships, then to be able almost always to confront with equanimity what life throws up would cohere with what one cares about. But if one's interests in closer interpersonal relationships are not so attenuated, then always to respond to circumstances with Spinozist detachment would be at odds with what one really wants. Free will calls for as much detachment, no more and no less, as will enable us to do and get what we really want.

This is not to say that we may often be less free than we would wish in being unable to modify or put aside recalcitrant desires, or that free will is not a matter of degree. But I see no reason to accept that most of us are entirely powerless in the face of recalcitrant desires. We can and often do manage to get beyond merely coping with disappointments and frustrations, so that over time we cease to be troubled by them.

It should not be counted as a limitation on our freedom, therefore, that most of us cannot match up to the Stoic/Spinozist ideal. One can imagine a non-human being of serene detachment who views our human commitment and passion with aloof disdain, but the judgement would be grounded in its being ***like that.*** We have no reason to suppose that calm detachment

is intrinsically desirable, without reference to human needs and interests, or that the idea that it is intrinsically desirable is anything other than a deception, typical of Spinoza's metaphysical idea of there being a god's eye view of the real nature of things.

To sum up, the ability to respond to any circumstance with equanimity and critical detachment calls for a high degree of self-determination, but self-determination is not necessarily bound up with this degree of calmness and aloofness. For most of us, such consistent disengagement from personal feelings and relationships is not an option. The challenge for us lies in marrying self-determination with normal human needs for love and recognition; in other words, to be or to become ***humanly free.***[32] To be humanly free is to be able critically to evaluate one's beliefs, attitudes, emotional responses, desires and values, and to be able to put that into practice, sometimes in being able to adjust one's aspirations to what is possible and worthwhile and sometimes even to stop wanting (or to want less) what one cannot have. It may be impossible for a prisoner entirely to adjust her wants to her circumstances, and to that extent she will be unfree, but if she is able to resist giving all her thought and energy to resentment and frustration she will have secured an important measure of freedom.

CONTROL BY THE PAST

In Chapter 1 I suggested that, in addition to interpersonal and moral attitudes, what fastens and reinforces the entrenched intuitions of the two camps in the traditional dispute about free will and determinism are various experiences (and the understandings these involve) of acting and deciding in various ways, which do not involve our being caused to act and decide as we do. These experiences can come into conflict with images or pictures of determinism and naturalism, among the shadows as it were, operating at a level below consciousness and language; conflicts that we struggle, sometimes blunderingly, to express. We are in the grip of pictures,[33] and pictures, being what they are, are not easily dissolved or refuted by propositions and arguments. One of incompatibilism's sustaining pictures (and one that is clearly linked to the Consequences argument) is that of determinism as binding us fast to the past, allowing nothing new or truly original to take place. The picture involves the past, together with the laws of nature, shaping the present from behind, so to speak; and always beyond the reach of our future-affecting actions.

This is a picture that compatibilists have laboured hard to dissolve,[34]

and the Stoic and compatibilist emphasis on the role of reason has gone a long way to doing so. The basic compatibilist argument is that providing we are able rationally to determine what it is that we really want and how to get it, and that we are able to act on that, we will be free to do so, and the fact (if it is a fact) that our actions are in some way shaped by the past is irrelevant to our being free.

The argument can be elaborated and made more persuasive as follows. The picture of determinism in which all our thoughts and actions are inescapably fused into causal chains stretching back to an infinity of points past is one in which those points that are our actions cannot be otherwise than they are. They cannot be avoided, cannot be done differently and cannot be undone. We seem to be unfree, therefore, in being unable to do otherwise than what is appointed for us at each of the links in the chains. We are unfree in being unable to do otherwise than we do. But if one confronts a choice between ϕ and 'I', and ϕ is what one really wants and thus the right and reasonable choice to make, whereas 'I' would satisfy an unwanted desire, the only free choice one can make is for ϕ, and the only reason for which one can freely choose it is that it is what one wants. If choosing freely consists in choosing ϕ because it is what one wants, then categorically to be able to choose ψ would not make one free or freer. ***To choose wrongly is not to choose freely.*** If you make the right choice you are free, if you make the wrong choice you are unfree. Free choice, as current jargon has it, is ***uni-directional*** or ***one-way rational.***[35]

As I said, this seems to me to go a long way towards undermining the picture of determinism in which we are bound by what happened in the past and the laws of nature and cannot do otherwise than they allow. We cannot, in this one sense, do other than we do, but we can still be free provided we can do what we really want because it is what we really want. Still, although this line of thought goes some way towards undermining the picture it does not quite dissolve it, and unless the picture is dissolved it can continue to cause determinism to be perceived as threatening even though we might explicitly reject any ***statement*** that this is so. If the picture remains undissolved, moreover, it will continue to generate intuitions that cause us (or some of us) to think that no matter how persuasive the contemporary compatibilist case, there is still a nagging sense of there being something it fails to account for: some troublesome fact or notion that it covers up.

It seems to me that the reason why the picture of control by the past resists the argument for unidirectionality of rational free will is that the rightness and reasonableness of a choice (given what one really wants) is taken to be just a contingent fact about it, which might not have been true

had some earlier state of affairs been different. In other words, what ultimately controls (what one will do) is how things were back then, together with the laws of nature, and the fact that what one chose or did turned out to be the right and reasonable thing to do was after all just a coincidental feature of what one chose or did, which it might not have possessed, and which in any case was brought about by what ***really*** determines how one acts or chooses. It was only superficially up to me that my choice was right and reasonable: really it was caused by the past and the laws of nature.

What is wrong with the picture, rendered explicit and understood in this way, is the notion that any choice or action that is right and reasonable just happens, coincidentally, to be so. It is largely a matter of coincidence that any of us is born and therefore that we have developed and been trained to have the capacity to reason and to act accordingly. But once any creature acquires a capacity to reason, and to be reasonable in action, the reasonableness of its choices is not just a matter of coincidence. From then on, until it dies or loses the capacity, reason will (to a greater or lesser extent) determine its choices. Reason, as the means of knowing what we really want and how to get it, is no coincidental feature of our choices, but rather a determining influence on them. The picture, therefore, needs not so much to be dissolved as repatterned. To say, without qualification, that we are determined by the past and the laws of nature leaves reason out of the picture. Reason constrains the shape and direction of the causal chains of which our choices are part; and the more it does so the freer one is. Reason can fashion what would otherwise be blind and blundering causal chains into the choices and actions of free agents.

Will exposing the mistaken assumption in the picture of deterministic control by the past succeed in dissolving or altering the picture? That will depend on whether others recognise it as having fuelled their intuitions and feelings about determinism, but I do not see how the picture can maintain its grip without the assumption. At any rate, it is not the only source of intuitive support for incompatibilism. One important prop of incompatibilist thinking continues to be the sense that we have a categorical ability to do or to choose otherwise than we do, and that our having such a capacity does not render out actions capricious or out of control. The argument that to choose freely requires that one chooses, unidirectionally, what is right and reasonable seems to ride roughshod over this intuition. A further, and some would say more important, source of commitment to incompatibilism is the notion of ***ultimacy:*** the idea that (some of) our actions and choices are 'up to us' and not owed to anything or to anyone else.[36] Again, the idea that free will consists in choosing or doing what we

really want because it is what we really want fails somehow to capture the notion that free actions and choices must be solely or ultimately our own doing. If it is possible to reconcile our prereflective attitudes and intuitions about free will, then, dissolving the picture of control by the past is not enough.

INCOMMENSURABLE CHOICES AND ULTIMACY

Not all the choices we face in life can be clearly settled by practical reasoning. There are 'Buridan's ass' choices in which the options that confront us can be equally promising or equally risky. And sometimes, in addition, choices can be ***incommensurable,*** which is to say, they will realise values and preferences that are so unalike that they cannot be brought into measurable equivalence. Choices such as these may even involve struggling to come to an understanding of what is important to us, in which our most fundamental values are brought into question; and therefore the criteria we would ordinarily bring to bear in engaging in practical reasoning might themselves be held up to scrutiny. Charles Taylor has argued that the ability to engage in radical re-evaluation of what is worthwhile is essential to our conception of self and responsibility.[37] Such re-evaluation is 'radical' in having no independent yardstick against which to carry out the re-evaluation, but just an inchoate sense of what is most essential to one's self, to which one tries to give descriptive expression. Being able to say what are worthwhile and what are unacceptable ways to act or live may sometimes be something we are unable to do with any confidence until we are directly confronted with having to make choices about them. For example, a woman may not want to have children, but might not really know until she becomes pregnant whether she would be prepared to go through with an abortion.[38] Differently, a person may judge that it is always better to 'go with the feeling' when it comes to important life decisions.

None of this exactly refutes the claim that the capacity to reason and reflect on one's desires and values, and the means of their realisation, is central to free will. Even people who always let their feelings have the last word in important decisions must evaluate and assess the alternatives to know what is at stake and what order of importance to assign to the decisions they must make. Nevertheless, if free will is identified with the ability to rationally evaluate one's values, aspirations, beliefs, desires and options, then it would seem to follow that where one's best option is not clearly indicated by reasoning - where it seems difficult or impossible to

determine what will best realise what one cares most about - a choice cannot be freely made. But to say that is to part company with what we would otherwise be strongly inclined to say about such choices: which is that they ***are*** freely made. We are faced again, therefore, with apparently strongly divergent intuitions about what kinds of acts and choices are to count as free.[39]

I think that we should be prepared to say that capricious choices are free only in the sense of not being forced on us by external circumstance. But this is not true of all choices in which reason is not decisive. To be free, as I have argued, is above all to be able to do what you really want because it is what you really want. Reason is central to free will because it is indispensable to knowing and getting what you want. Nevertheless, ratiocination has its limits. It is sometimes true that what one really wants is most clearly revealed in a feeling about what to do, rather than by careful self-analysis. And if there can be great difficulty in giving faithful expression to an inchoate sense of what really matters, a sense of where one's feelings are pointing (since feelings are much at issue) may be a more reliable guide than hard-headed reasoning. One's feelings can after all be reasonable.[40]

This assumes, however, that there is always one thing one most wants, rather than two or more incomrnensurables that one equally wants. But if a choice cannot be settled by reasoning through the pros and cons, and if as it happens there is no sense in which an option best realises what one most wants, in what sense can the choice be free? We can say that a person is free in being able to come to a clear appreciation of the options she faces and what they imply, but if there is nothing to choose between the options in terms of satisfying her wants, there is, it seems, no way she can choose freely. It does not follow from this that she must choose capriciously. She might attempt to let her feelings decide in the hope that, after all, one option will best realise what she most wants (she might not know for sure that there is nothing to choose between her opposing wants). Although this will save her choice from capriciousness, however, it will not make it free.

The way out of the problem, I think, is to remember that freedom is a contrastive concept. A choice is free if it is not unfree. If one has reflected carefully and accurately about a choice one faces, and there is nothing to choose in terms of what one really wants, one may be forced to resort to some form of tie-breaking device, but provided one is not prevented from choosing what one really wants by some motivational or external impediment, then the choice can be free. A person chooses unfreely if she chooses what she does not really want to choose, and freely if she chooses what

she does want. If there is nothing to choose, in terms of what she really wants, between evenly matched or incommensurable options, she will be free in choosing either way.

I think that this is right, as far as it goes; and it gets us over the theoretical problem about whether options that satisfy equally matched or incommensurable wants can ever be free. But the response fails to account for a strong sense in which, when we have to choose between incommensurables or evenly matched options, the choice is 'up to us' or 'up to me': what I referred to above as the sense of ***ultimacy.*** This sense of ***up-to-me-ness,*** moreover, seems also to be involved in our experience, prior to making a choice or decision, of being able categorically to choose or decide either way.

What is involved in the sense of ***up-to-me-ness*** we feel when choosing between incommensurables, however, is not just the freedom to choose either way, but also a sense of ***agency.*** The ***up-to-me-ness*** we feel in being free to choose either way is itself a sense of agency, and one that is distinguishable from the sense of being able to choose either way and is a response to it. There is probably no precise distinction between conceptions of agency and free will in our prephilosophical intuitions about choosing and acting, but lack of a clear prephilosophical sense of the distinction does not imply that any hope of arriving at a coherent and exact philosophical understanding of agency and free will is forlorn. One reason ***freedom*** serves as a prephilosophical label both for what is properly so called and also for agency is that one's sense of agency in respect of many choices ***depends*** on the sense of being free to choose. But since there is no contradiction in the idea of an unfree action,[41] the conceptual distinction between agency and free will is far from being a theoretical invention. Provided our philosophical understanding can account for the respective prephilosophical intuitions without doing violence to them, that our prephilosophical ***labelling*** of intuitions turns out to be theoretically inconsistent is of no great significance.

The concept of agency is logically prior to that of free will: we cannot begin to talk of free will without some conception of that which is free. The sense of ***up-to-me-ness*** we face in choosing between incommensurables or evenly matched options is not, as such, a sense of being able to choose either way, since this is a sense we can and do have even where the rational course is clear. It is instead a sense of being forced back on oneself. When reason offers no guidance about what I should choose, and when perhaps there is not even an expectation about how I will choose, I am left with the bare necessity of choosing and with my own capacity for choice. When there are good reasons for choosing as we do, we tend

to choose with little thought to the fact that it is us and not the reasons doing the choosing. Having to choose without reasons brings us face to face with our own agency.[42]

Incompatibilists such as Robert Kane will no doubt object to this that what I have said about the sense of ***up-to-me-ness*** runs together the issue of our being able to choose or act otherwise than we do with that of ***ultimacy*** or ***ultimate responsibility***, and that in any case what I have described as ultimacy falls short of what incompatibilists have taken it to be. According to Kane the idea of ultimacy is that of 'buck-stopping' responsibility, such that to be ultimately responsible for an event or state of affairs an agent must have causally contributed to its coming about and nothing and no one other than the agent could have been a sufficient condition for her so contributing.[43] Ultimate responsibility is worth wanting, Kane argues, because it is needed, ***inter alia***, for moral responsibility, creativity, dignity, individuality and being able to say that one acts of one's own free will.[44]

The sources of the incompatibilist idea of ultimacy seem therefore to be diverse. To begin with, however, I think that we can eliminate moral responsibility from the running, since, for the reasons I gave in the previous chapter, I see no way in which the addition of the idea of ultimacy (however it is understood), or of the causal indeterminacy that Kane takes to be required for it, could be shown to make the reactive attitudes and related practices any more fair or justified than they would be without it; and it is the fairness of these attitudes and practices, as I have argued, that is at the root of our worries about their relation to determinism.

Moving on to dignity and individuality: it is true that our hopes and desires in respect of these goods can be threatened or troubled by the idea that what we do and say and think may be derivative or unoriginal, but it does not follow from this that they are related to any idea of ultimacy (or at least not one that can be explicated in terms of causally sufficient conditions and so forth). When we are troubled about our dignity or individuality, it is always and without exception, tacitly or explicitly, a worry about how we stand in relation to others (e.g. 'Am I just going along with what others expect of me?', 'Does he think I'm stupid?', 'Was it really me who thought of that?', 'I should have stood up to him' and so on). It is possible to think that there must be an ultimate basis to individuality and dignity, and to imagine that this is ruled out by determinism, but this, it seems to me, is simply to read philosophical preoccupations into what we would ordinarily say or think or feel about individuality and dignity.

Those who have speculated about the sources of creativity do not seem, on the whole, to have been especially concerned that it should be thought

to reside within the creative person in some ultimate sense. Often enough it is suggested that creative people are divinely inspired (Mozart being an obvious example), or that they are able to 'tap into the cosmos'. Contemporary pop-explanations of creativity are apt to be couched in terms of angst, obsessiveness, depression and neuro-chemical transmitters. And none of these explanations have been taken to imply that creativity is any less valuable or less real.

It is possible that philosophical reflection can lead us to associate the idea of ultimacy with all manner of other concepts and images, but I think that its real prephilosophical basis is in the experience of confronting alternatives where no choice or course of action is clearly indicated by practical reasoning and we are sharply confronted by the bare need to choose and the bare capacity we have as agents to make choices.[45] Thus, while the notion of ultimacy can be distinguished from our apparent capacity categorically to choose or do oiherwise, and while different considerations may determine whether either notion is illusory, the sense of ultimacy or ***up-to-me-ness*** attaches to experiences of facing choices and alternatives for action where reasons offer no unambiguous guidance about what to do.

One source, therefore, of the problems philosophers have faced in reconciling rationality-based accounts of free will with intuitions about ultimacy and 'the liberty of indifference'[46] has been a failure clearly to distinguish between intuitions and experiences having to do with agency and those relating to free will. Our sense of agency requires that when we have a choice it is we who do the choosing, and this in tum requires that we understand ourselves as having an identity distinct from any of the criteria we bring to bear in making a choice. Desires do not choose or act, and neither do hopes or fears or reasons. Thus, while it may be that one of the options about which I have a choice will clearly best realise what I really want, and that choosing freely means choosing that way, if I do not have the capacity, in some as yet unspecified sense, to choose both reasonably and umeasonably, it will not be ***me*** who does the choosing; and if I do not understand myself as having the capacity to choose either way, I cannot understand myself as doing the choosing. If my experience or understanding is not one in which I could, although I don't, choose reasonably, then I will experience some craving or fear or base desire as doing the choosing. If my experience is not one in which I could, although I don't, choose unreasonably, I will experience my best reasons as doing the choosing.

The criticism I made earlier of the Spinozist/pure rational ego paradigm of freedom was that it takes one possible way of being free as exhausting

all of the possibilities. What we should be interested in, as I argued, is what we need in order to be ***humanly free.*** It is possible to imagine a very advanced acting and deciding machine, which is sentient and 'aware' (in a strong sense) of its acts and decisions and the reasons for them. If its acts and decisions follow from good reasons as conclusions follow from premises, as the idea of one-way rationality might be taken to imply, we could say that the machine acts and decides freely. But it will not act and decide freely in the way that we do, since it will lack any sense of its acts and decisions as owed to its own agency. Lacking a sense of agency, there will be a dimension of judgements, attitudes, questions and feelings, about what it wants for itself and how it wants to be, that will be entirely beyond it.

Consider Galen Strawson's imaginary race of 'natural Epictetans', who always get what they want, are never undecided and never have to struggle to make a choice, because of a capacity to effortlessly adjust their wants to what circumstances will permit.[47] Although in being free to choose and always getting what they want, we could say that they are always free, notions such as freedom and choice would have no positive meaning for them because they would never know what it is to be without freedom. Strawson concludes that although they satisfy objective conditions for being described as free, in having no sense of being able to choose otherwise than they do they do not really have freedom.

It is true that since freedom is a contrastive concept, creatures who are never denied what they want would have no understanding of freedom-talk. But what they would lack is a sense of agency rather than freedom. They would be free but in a different way from us; or, as we might say, they would lack the kind of freedom we have, rather than freedom as such. We might be inclined to see ourselves as freer than the natural Epictetans, rather than just free in a different way from them, but all that would mean is that there are things we can do that they cannot. By the same token we are less free than birds in being unable to fly unsupported and less free than fish in not being able to live under water. If the freedom of the natural Epictetans is not what most of us have in mind when we think of ***freedom,*** then that is because we are apt to think of freedom in the way that ***we*** have it, which is to say: ***free agency.***

There is no divergence, therefore, in our prereflective intuitions between the 'liberty of spontaneity' and the 'liberty of indifference'. The experience of being able to do otherwise than we do, or of being able to choose categorically either way, does not contain a divergent intuition about free will from that of being able to do what we really want: it is a feature of our sense of agency. A sense of agency requires a space between rational

judgement and the Will. Our sense of agency is also the root of the experience of ***up-to-me-ness*** we have in respect of incommensurable or evenly matched choices.

What remains to be explained is whether we are agents such as we take ourselves to be, and whether the experience of being able to act otherwise than we do, and of its being ***up-to-us,*** is veridical or illusory.

CONCLUSION

In this chapter I set out to examine what is required for an account of free will as being able to do what we really want because it is what we really want, and to consider whether this idea is at odds with other prereflective intuitions and attitudes we have about free will. I argued that free will, thus understood, calls for qualities such as intellectual organisation, critical detachment, and proper motivation, which have famously been associated with the Stoic tradition of thinking about freedom. I argued that the idea of free will as thinking and acting in accordance with what is right and reasonable does not imply an external or objectivist standard of morality (although it does not exclude it), which might conflict with what we really want; but that our natures place constraints on what we can count as important and valuable and, consequently, on what we can want. I argued that we can be prevented from doing what we want either by 'external' or 'internal' circumstances and that this is explicated in the distinction between freedom of action and free will. I defended the Stoic idea of adapting our wants to what we cannot change against the charge that this is beyond our human capacities, while agreeing that the standard Stoic conception is too much at odds with human nature and interests. I argued that free will should not be identified exclusively with detachment from interpersonal needs and attitudes, or with self-sufficiency: that it is possible to be humanly free. I considered the way in which incompatibilist intuitions about determinism and free will might be fuelled by a picture of determinism binding us to the past and showed how the picture could be dissolved or repatterned by thinking about the key role played by reason in Stoic and compatibilist accounts of free will. Finally, I considered the objection that our intuitions about incommensurable choices are at odds with the claim that free will consists in being able to do what one wants because it is what one wants. I argued that our intuitions about incommensurable choices are not about free will, as such, but about our experience or sense of agency, and that with an appreciation of the distinction there

is no reason to suppose that our prephilosophical intuitions about free will are radically divergent.

If our prereflective intuitions, and the experiences in which they are grounded, are not radically divergent, that is not enough to show them to be veridical. In the following chapter I consider whether causal determination of decisions would have the consequence that the way we experience deciding, including the sense of being able categorically to decide either way, is illusory.

4 Can We Experience our Decisions as Caused?

I have argued that the sense of ***up-to-me-ness*** or ultimacy we have in respect of choices and actions where reasons give no clear guidance about what to do, and where it seems that we can choose or act either way, is properly understood as a sense of agency rather than of free will. For the purposes of this chapter, however, the distinction is unimportant and, in the interests of avoiding a tedious insistence on a point that is irrelevant to the issues I will be discussing, the freedom I will have in mind when referring to the ***experience of freedom*** will be the freedom that incompatibilists take to be threatened by determinism: that of being able to choose or act either way, where it is the agent and nothing and no one else that is thought to do the choosing. The acts and choices to which the sense of ultimacy is thought to attach are those which call for deliberation and decision, and it is with whether our experiences of deliberating and deciding are veridical that this chapter is concerned.

Incompatibilists argue that if determinism is true, then the experience that some decisions are truly and inescapably up to the agent is false, so that no matter how strongly it may appear that I have to 'come up with' a decision, the content of that decision is already fixed.[1] Even if some incompatibilists concede my arguments about control by the past and moral responsibility, they may still insist that the contents of our experiences of choosing or deciding freely are incompatible with determinism (and therefore incompatible with belief in determinism).[2] Those incompatibilists who believe determinism to be true, therefore, are committed to the view that no matter how pressing and how real the burden of responsibility may feel in respect of some decisions, this is not how things really are.

The claim that if determinism is true our experiences of being free to deliberate and decide are illusory, implies not only that there is something amiss in such experiences, but also that there are logically possible veridical experiences of our deliberations and decisions. Such experiences, although logically possible, might not actually be possible for us, perhaps because our physical or cognitive natures place them beyond our comprehension: in which case we would labour under a necessary illusion. Such experiences must, however, be at least logically possible if the claim of illusoriness in respect of our actual experiences is to have any meaning;

for what could ***illusory*** mean except in contrast to what is in some sense ***veridical?***

If determinism is true, incompatibilists claim, then my experiences of it being entirely up to me how I decide are false because any decision I make will be a consequence of causally sufficient antecedents and therefore I could not have decided otherwise. For present purposes it does not matter whether the incompatibilist argument is sound: it is sufficient to note that it commits them to the claim that a veridical experience of a causally determined decision would have to include the decision's ***being causally determined*** and therefore unavoidable.

In this chapter I will examine in what ways and in what circumstances we might experience decisions as caused, and whether such experiences would be incompatible with our also experiencing such decisions as free. Before considering whether it is possible to experience decisions as causally determined, a few remarks are in order about what in general it is to experience events as caused. To experience decisions as caused is to apprehend them as falling under a general relation of causation. Whether it is possible to experience decisions as caused will depend, therefore, on what it is in general to experience one thing as being caused by another.

THE EXPERIENCE OF CAUSATION

To experience an event *C* as causing another event *E* is to experience *C* as singularly related to *E* such that when *C* happens one expects a type *E* event to follow; and that this expectation is based on an understanding that, ***ceteris paribus, C*** is required for E.[3]

This analysis of the experience of causation has some similarity to Hume's second definition of cause as necessary connection, according to which the cause is an 'object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other'.[4] Both Hume's definition of causation and my summary of the experience of causation assert that, in addition to the perception of the events related as cause and effect, there is an identification of the cause as being of a certain type and an expectation of the effect. And since the identification and expectation are not necessarily connected to one's perceptions of the events, both accounts allow that we can experience events to be causally related which in fact are not, and that we fail to experience events as cause and effect which in fact are.

The analysis differs, however, from Hume's definition of causation in

four significant respects. (1) It makes no reference to contiguity. And properly so, since it is possible to experience an object as acting at a distance on another: the effects of magnets being an obvious example.

(2) It claims that the expectation of the effect is based on an understanding that, ***ceteris paribus,*** the cause is required for the effect. An example should serve to prove the point. If I set my alarm clock for a minute before sunrise, when it goes off I will have an expectation that the sun will rise in a minute, and that expectation will have been produced by the alarm. But I would not experience the sunrise as an effect of the alarm. I will not experience it as such because I will know that even if the alarm had not gone off, the sunrise would still have happened.[5] Not all expectations of events, therefore, count towards experiencing those events as caused by events that give rise to the expectations. We might say then that the expectation we have that an effect will follow a cause is a ***causal expectation,*** where this indicates that the expectation is based on identification of the cause as being of a certain event type together with the belief that, ***ceteris paribus,*** the cause is required for the effect.

(3) The analysis asserts that to experience one event as causing another involves experiencing that event as singularly related to its effect. The word ***singularly*** signifies that where, in addition to what we call the cause of an event, there are several ***background conditions*** that are jointly required for the occurrence of the effect and which, taken together with the cause-event, make up what philosophers call the 'causal circumstance' for it, it is the singular event alone, and not the whole causal circumstance, that is experienced as the cause. Hume's definition does speak of ***'an object*** precedent and contiguous to another . . .' (my emphasis), but there is no indication that he intends this to rule out background conditions being experienced as part of the cause. In addition it is Hume's first definition of cause as 'An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedency and contiguity to those objects, that resemble the latter',[6] which initiated the tradition of regularity analyses that treat cause and effect as instances of invariant relationships between causal circumstances and effects.

That it is a singular event rather than an entire causal circumstance that we experience as the cause of an event contradicts Hume's claim in the first definition that part of what we mean by cause is grounded in the experience of regular relationships between event types. The objection that it is possible that a unique event be the cause of another (the effects of detonation of the first atomic bomb, for example) is well known. To experience one event as causing another, even though one has never experienced

an event of that type before, it is sufficient only that one has what one takes to be good enough reason to believe that the effect will follow (say because one has theoretical knowledge, or that one has it on good advice). In addition, it is quite possible to experience events as causally connected on the basis of only one previous experience of their occurring together, because, for example, the effect is sufficiently striking. A child who grasps a nettle does not need to have the experience repeated to appreciate the relationship between grasping the nettle and feeling pain as one between cause and effect. In the first case, however, she will not have an occurrent experience of grasping the nettle as causing her pain, but rather two distinct experiences that she will reflectively judge to be related as cause and effect.

My claim that to experience an event as causing another is to experience ***just that*** event as doing the causing (rather than the whole causal circumstance), although consonant with 'singularist' theories of causation,7 does not imply that background conditions play no part in the experience of causation. Background conditions enter into the experience of causation in being required for the identification of the cause and for the causal expectation of the effect. An event that is said to be the cause of another, or is experienced as causing another, can only be conceived as doing so on the understanding that the background conditions that enable it to do so are in place. If a requisite background condition is missing, the so-called cause cannot be the cause it would otherwise be. An event is therefore properly identified as belonging to a particular causal type only if the requisite background conditions are in place. Without those conditions the 'same' event will fail to be an event of that type, which is to say, the background conditions enter into its being an event of a particular causal type. Thus in expecting an effect to follow from an event one has identified as being of a particular causal type, one does so on the assumption that the requisite background conditions are in place. For this reason it might be thought that we ***do*** experience the entire causal circumstance as causing the effect. That we do not can be seen by thinking about how we experience cause and effect in respect of switching on an electrical appliance such as a radio. Under normal circumstances one will expect to hear broadcast sounds when one pushes the on-button. One would not have the same expectation if one pushed the button knowing that the radio was not plugged in or that there was a power cut. Nevertheless, assuming the radio is plugged in and the power is on, it is one's pushing the on-button that one experiences as the cause of the radio coming on and not ***one's pushing the button conjoined with the thought that the radio is plugged in and the current is on.***[8]

In respect of the experience of causation, therefore, background conditions are exactly that: ***background*** conditions, which is to say, conditions that are in the background of the experience, or, rather, conditions the thought of which is in the background of the experience. The difference between background conditions and the event that is experienced as occurrent cause consists in their being, experientially, in the background and the cause being in the foreground.[9]

(4) The idea of natural necessity does not form part of the analysis. According to Hume's account of causal expectation, the idea of the cause determines the mind to form the idea of the effect, and the impression of the cause to form a more lively idea of the effect. Critics of Hume have pointed out that his supposed definition of causation actually employs a range of causal terms whose meaning it is supposed to explain.[10] A further objection is that the definition fails to explain the idea of natural necessity. Even allowing that natural necessity is a different concept from that of logical necessity, it does have this much in common with it: that we use it to refer to ***what cannot be otherwise.*** Even if no more sense can be given to the concept than reciting these words, at a bare minimum it must mean that. But if my mind is determined by the impression of one event to 'form the idea' of another, or even 'to form a more lively idea of it', it may still be possible for me to realistically entertain the possibility, however slight, that the event I am having a lively idea of will not happen; where 'realistically entertain the possibility' means really believing that it might not happen and not just imagining, as a logical possibility, that it might not happen. If I can realistically entertain the possibility that an event might not happen, then I cannot at the same time regard it as something that ***cannot be otherwise.*** There are occasions when as a matter of fact people do find it psychologically impossible to realistically entertain the possibility that an expected effect will not take place. Someone who sees a person fall from the roof of a fifty-storey building, for example, might well find it impossible to believe that the man hurtling to the ground might not be badly hurt. In that case we would rightly describe her experience of the relationship between fall and landing as involving the idea of natural necessity.

But not all experiences of causation are like that. It is possible to form a causal expectation, on encountering an event one identifies as being of a particular causal type, without believing that it is impossible that the expected effect should not happen. When I press a button on my radio I expect sound to come from the speakers and that is a causal expectation, but the expectation does not render it impossible for me to believe that there might be no sound. Whether or not I experience an effect as necessitated

by its cause will depend on how likely I think it is that failure of a requisite background condition might occur. The person watching someone fall from a great height will experience the sequence as necessary because they will believe that the possible non-occurrence of any requisite condition would be incredible. Therefore, even if Hume's account had captured what it is to have a causal expectation, it fails to explicate what it is to experience an event as physically necessitated. And this is unsurprising, since experiencing an event as caused does not require that it be experienced as necessitated.[11]

There is nevertheless a conceptual link between the experience of causation and the idea of natural necessity. As I have argued, to identify an event as being of a particular causal type, and to have a causal expectation of an effect, requires the assumption that the requisite background conditions are in place. As I have also argued, this assumption is compatible with the further belief that a background condition might not be in place. Indeed the assumption and the belief can be synthesised as one belief that the background conditions are probably in place. Since identification of the singular cause and the causal expectation that goes with it are based on the assumption that the requisite background conditions are in place, and since any belief that what is expected might not happen derives from the belief that one or more of the background conditions might not be in place or might not occur (or that the cause may not have been accurately identified), anyone who entertains a causal expectation has a belief whose content entails the proposition 'if all of the requisite background conditions are in place and the cause occurs, the effect ***cannot fail to happen***': that it is causally necessitated. Without implicit belief in this proposition, there could be no causal expectation and therefore no experience of causation. This being so, any philosopher who argues that it is consistent with our idea of causation that where one event causes another it is possible that exactly the same event and background conditions should occur and the effect fail to happen, has claimed what they cannot consistently believe.[12]

Having argued that belief in causal necessity is implied by the experience of causation, I emphasise again: an event that is experienced as caused need not be experienced as causally necessitated; although some are. What ***is*** required is that one has a causal expectation of the effect: a condition that clearly cannot be satisfied if one believes that the so-called effect might just as well not occur.

The experience of causation is complex. A more thorough and nuanced account than mine might reveal much more complexity. It seems to me, nevertheless, that the analysis is right in its essentials, and particularly in respect of causal necessitation.

CAUSATION AND DECISIONS

If this is how we experience one event as causing another, is it possible for an agent to experience this in respect of her decisions? Several philosophers have argued that it is a logical or conceptual impossibility for a person to know a decision before she makes it.[13] According to these philosophers a caused decision would be one that could, in principle, be known before it is made by identifying the antecedent causal circumstances for it and applying the relevant law statements that connect the circumstances to the decision (or a reliable inductive generalisation that does so), thus enabling the decision to be predicted. This, it is argued, is a conceptual impossibility, since prior to making a decision one must be uncertain about what it will be, which is incompatible with knowing what it will be.

Against this it might be argued that we can experience our actions and decisions as caused, and often do. Those soft determinists who follow Hume in believing not only that freedom and moral responsibility are compatible with determinism but that they positively require it could claim that actions and decisions are sometimes caused by stable virtues of character, and experienced as such.[14] Let us examine this claim more closely. Suppose a person holding public office is regularly offered bribes, which she unfailingly refuses. She may feel content that her character is of such firm and incorruptible stuff that she will always refuse a bribe. Is the unvarying nature of her responses and her confidence in their continuation enough for her to experience her refusals as caused? Is it enough to satisfy our analysis of what it is to experience causation? To answer that, we would need to know in the first place what her refusals are caused by (and what she experiences them as caused by). There is an unvarying relationship between her being enticed and her refusing, but if her refusals were caused by the events of her being offered bribes, they would fail to count as decisions. Those who argue that freedom and moral responsibility positively require determinism, as I understand it, think of free actions as those caused by character rather than by external stimuli.

A person's character is not an event or an object, and therefore, according to our analysis, not something that could be experienced as a cause. Moreover, one's idea of one's character is inferred from one's actions, moods, dispositions and so on, and is not something we perceive as singular and unified. Therefore it could not satisfy the requirement that what is experienced as a cause be a singular event. The singular component of character that a Humean would presumably take as causing the official's refusal of bribes would be her strong conviction that bribes should not be accepted.[15] It is possible to have a strong conviction about something

without having to have an occurrent conscious experience of that conviction, but only if the conviction is occurrently experienced (or only if one's sense of it is) could one experience it as causing anything.

Say then that the official experiences her refusals as regularly following not only the attempted bribe, but this coupled with her strong conviction that bribes should not be accepted. **If** the conviction were so strong that she could not resist it no matter how much she might want to, then she could not experience her refusals as decisions she had made. **If,** on the other hand, she experiences a real struggle whenever she is offered a bribe, even though she has always eventually refused, then, logically, just in so far as it is a struggle, she experiences it as a real possibility that she will accept the bribe and therefore she will lack the causal expectation that would allow her to experience her strong conviction as a cause of her refusal to accept the bribe.

On the other hand, perhaps she never experiences inner opposition to her strong conviction. It would not do to object to this that, since she will not have precise descriptions of her mental states, she can never know for sure that her strong conviction and not something else is the cause of her refusal, or, for the same reason, that she could not know for sure that her refusal would actually take place. We never know this much about many everyday causes: instead, we assume that the usual background conditions are in place and that things are as they seem. In that case I think it could be claimed that the official experiences her refusals as caused.

Someone who experiences her decisions as caused in this way would not be in the position of knowing her decisions ***before*** she made them. **If** she experiences her decisions as following effortlessly from being confronted with a bribe, together with her strong convictions, the decision to refuse the bribe might be thought of as occurring simultaneously with the expectation of it, which is to say, as soon as the expectation is formed it is realised. Incompatibilists might argue either that this would not involve a genuine decision or that it would not involve genuine causal expectation, but I can see no obvious reason for accepting this.

It might be argued instead that a so-called decision which is experienced as occurring unfailingly and effortlessly would, like any effect, not be experienced as requiring anything on the part of the experiencer. A 'decision' I experience as caused would on this account be rather like a hiccup or a sneeze: something that happens to me rather than something that requires my participation or has its source in me ***qua*** agent.

Another argument would be that causal expectation of effect would be obstructed in the case of a genuine decision (or what is genuinely experienced as a decision), by its connection with reasons. Decisions are

experienced as following from reasons just in the sense that they are carried out ***for*** particular reasons. Davidson has shown how reasons can be thought of as causes of actions,[16] but it should still be accepted, even if nothing else follows from it, that the reasons an agent has for a particular action need not rule out, all by themselves, her failing to perform that action. In that case, if a decision follows from reasons that do not render it as inevitable or unfailing, it might be said to occlude or obscure (just as the Sun obscures the stars) any possibility of causally expecting the effect.

Both of these objections seem to me to beg the question against the Humean argument that it is possible to experience decisions as caused in virtue of their unfailing connection with stable virtues of character. While it may be true that the connection between reasons and actions in general is not one in which occurrent awareness of reasons leads to a causal expectation of actions, this does not rule out the possibility that the kind of person we have been considering would experience her refusal of bribes as following inevitably from her opposition to corruption. And to say that a decision which is experienced as caused would be one in which the experiencer cannot feel that anything is required from them invites the rejoinder that while such may be the case in respect of involuntary behaviour, or events that are external to a person, there is no reason to accept that this is a necessary feature of all effects. (Nevertheless, the argument does seem to me to have some substance and I will return to it later on.)

Incompatibilists might object that the official's 'decision' is not a genuine decision,[17] and at any rate that cases such as this do not involve the experience of deciding freely in the strong and important sense they take to be threatened by determinism. Galen Strawson, for example, identifies ***difficult*** and ***painful*** choices as 'the central fact of the phenomenology of freedom':[18] choices that involve real struggle in reaching a decision. If the Humean example of a person who experiences her decisions as caused is allowed to stand, is it possible that a person experiences decisions that involve real deliberation and difficulty as caused?

Those incompatibilists who have thought that it is a logical impossibility that one might know what one will decide before one does so argue that making a decision involves a passage from the uncertainty of deliberation to intentional certainty about what one will do and that being uncertain about what one will do is logically incompatible with knowing what one will do. Both Ginet and Taylor have argued that if any decision is caused, it would be logically possible for the decider to know that it will happen, provided he knows the relevant circumstances and causal laws connecting the circumstances with that decision.[19] For this reason, they

argue, it is impossible for a decision to be caused: either decisions exist and are not caused, or decisions, as we ordinarily understand and experience them, are illusory.[20] Even if it is allowed that we can meaningfully speak of decisions that are not preceded by uncertainty, as would be the case with the virtuous official, where decisions do follow on from any degree of uncertainty and deliberation, or, as we should say, where we ***experience*** decisions as following on from deliberation and uncertainty, it is an impossibility, according to incompatibilists, that we should also experience them as caused.

Before proceeding further, then, I want to look more closely at the argument that it is impossible to know a decision before one makes it, and also at the characterisation of decisions and deciding on which this argument rests.

DELIBERATING, DECIDING AND INTENDING

To make a decision or to 'make up one's mind', according to Ginet, is to become sure about what one will do, and, indeed, that coming to a decision ***just is*** coming to a state of certainty about what one will do: 'the whole point of making up one's mind is to pass from uncertainty to a kind of knowledge about what one will do or try to do'.[21] Hampshire and Hart likewise argue that while a person is

> making [a] decision, and while he is reviewing reasons for acting in one way rather than another, he must be in a state of uncertainty about what he is going to do. The certainty comes at the moment of decision, ***and indeed constitutes the decision,*** when the certainty is arrived at in this way, as a result of considering reasons, and not as a result of considering evidence.[22] ***(My emphasis)***

Hampshire and Hart rightly identify the making of a decision as involving transition from a state of uncertainty to one of certainty, and infer from this, as does Ginet, that deliberation and decision are cognitive states. This can be seen in Ginet's claim that certainty about what one will do is the ***whole point of making up one's mind,*** and in Hampshire's and Hart's suggestion that the coming of certainty constitutes the decision. What distinguishes the termination of uncertainty about what will happen by a decision, according to Hampshire and Hart, is that unlike a prediction, which is inferred from evidence, a decision follows from reasons for acting in one way rather than another.[23] To the extent that a person's actions

are the result of a decision, they argue, rather than something that has been predicted, we call them voluntary.

Although Richard Taylor does not regard deciding as a cognitive state, he does agree with Ginet, and Hampshire and Hart that deciding resolves doubt about what one will do, such that it is impossible to try to decide by deliberating while knowing what one will decide.[24] O'Shaughnessy, who is also careful to distinguish deciding as the 'onset of practical commitment' rather than that of cognitive commitment, views deciding, likewise, as a 'coming-to-intend', leading from practical uncertainty to the resolving of doubt.[25]

I want to argue that these philosophers have misunderstood the way in which deciding resolves uncertainty; that this has led them to mischaracterise what it is to decide and how deciding stands in relation to deliberation, uncertainty and intention; and that this mischaracterisation, involving the idea that deciding is, or that it entails, a cognitive state, has been central to the argument that our experiences of deciding are incompatible with causal determination of decisions. Once the real nature of deciding is identified, it can no longer be seen as a cognitive state, or even as a quasi-cognitive state.

A decision is an act of resolving deliberation by a ***marking of intention*** to act. Deliberation about what to do is the activity of attempting to determine what to do by reasoning between alternatives. Other than in a trivial sense, therefore, the aim of deliberating about what to do is not to arrive at a decision but to arrive at a conclusion about what to do. The uncertainty that is resolved by making a decision is uncertainty about ***what to do*** and not uncertainty about ***what one will do.*** Uncertainty about what one will do is a cognitive state, but uncertainty about what to do is not. A decision resolves uncertainty about what to do, and brings the process of deliberation to a conclusion, by affirming that this is what ***will*** be done.

In saying that a decision is a marking of intention to act I do not mean to imply that decisions must necessarily involve speech acts, or indeed that in making a decision one must at least mentally say what it is that one intends to do. Decisions can be and are made privately and wordlessly (even in a mental sense). But to make a decision one must understand oneself to have brought deliberation to a conclusion, and to have done so by selecting the preferred alternative as what one will do. That this is so can be seen from the fact that it is sometimes possible to decide simply by uttering words. This happens in situations where one is called upon to

decide what is to be done, on matters of organisational policy, perhaps, or just where to go to eat. In such decisions, where the decision is ours, we decide by simply making a declaration. If you have left it to me to say where we shall go for breakfast and I say 'okay, then we'll go to Delancey's', then by that alone I have decided: my declaration has, as we might say, ***illocutionary force.***[26] My deciding that we will go to Delancey's in this way is logically compatible with our not going there because you have thought of somewhere better, or because I didn't really want to go to Delancey's and only said it in order to see what your reaction would be. Any of this may be true, without altering the fact that I made a decision just by uttering the words 'we'll go to Delancey's'. Clearly, then, it is possible to decide without being sure what it is that one will do. All that is required for making a decision is that the process of considering alternatives that we call deliberation is brought to a conclusion by selecting from among them.

Those who take deciding to be a passage to a state of certainty about what one will do might object that this is a special sense of deciding, which differs from deciding for oneself in at least one fundamental respect: that one can publicly decide without intending to do anything, but one cannot decide for oneself without at the same time forming an intention to act on one's decision. This is true, but just for the reason that if one definitely intends to do something, one will usually know that it is one's intention.[27] It would not be possible to resolve deliberation by marking an intent to do something one knows one does not intend to do. Thus, if I lack the intention of doing something, then for as long as I lack the intention, I cannot understand myself as deciding to do it. For the same reason, one cannot deliberate about doing something, other than in an attenuated hypothetical sense, if one knows that one would never do it. When it comes to public decisions, since I can make decisions for others, what I decide may or may not depend on what I will do and therefore may or may not depend on my intentions. But if it is known by me and those to whom the decision is announced that what I have decided on will not happen, either because it cannot happen or because I announce at the same time that I will not comply with my 'decision', I could not meaningfully be regarded as having made a decision.

Is this just hairsplitting? I have argued that decision is the resolution of deliberation by the marking of an intention to act, and not a mere passage from uncertainty to certainty about what one will do. But those I am arguing against claim that the making of a decision is a passage to certainty about what one will do ***by the formation of an intention.*** Surely if I agree that one cannot decide without intending to do what one has

decided (or with public decisions, declaring *that* it will be done), then couldn't it be said that my disagreement is a trivial insistence on a particular description that turns out to cover the same points of substance as the description to which I am objecting? My opponents would agree with me that decision terminates deliberation and I would agree with them both that it brings to an end uncertainty about what to do and that it requires the formation of an intention.

This apparent agreement, however, conceals substantial disagreement. In the first place, and as I have already mentioned, the uncertainty that is terminated by the making of a decision is uncertainty about what *to* do and not about what one *will* do. Those who regard decision as a passage to certainty mistakenly treat these as equivalent, or that the former implies the latter. Ginet, and Hampshire and Hart, who believe that decision making is a cognitive state, make the mistake of treating certainty about what to do as equivalent to certainty about what one will do, whereas Taylor and O'Shaughnessy believe that certainty about what to do, although not equivalent to certainty about what orie will do, does imply it. Therefore they are all committed to claiming that loss of certainty about what one will do must be accompanied by loss of certainty about what to do. I contend that it is only at the moment of making a decision that it is necessary that one intends to do what one has decided and thus that one believes one will do what one has decided. After the decision has been made one's intention may waver, and one will therefore be uncertain what one will do. Unless one goes back to deliberating, however, one's decision about what to do will be unaffected by one's subsequent lack of intention. One can therefore be certain about what *to* do while being quite uncertain about whether one *will* do it. Lack of intention to do something is an obstacle to deciding to do it, but not to *having decided* (in the sense in which one's decision still stands).

It is certainly possible that a wavering intention can lead one to review a decision and, thereby, cause one to begin deliberating again. My lack of intention might signal to me that what I have decided conflicts with my values in a way that I had not fully appreciated when I made the decision, or that I lack what it takes to carry the decision out. This would be sufficient grounds for reviewing the decision, and possibly for changing it. On the other hand what causes my lack of intention might be something that I specifically rejected or repudiated in making my decision; something I wished to overcome, perhaps, like laziness. In that case my lack of resolve would not lead me to change my decision and might even lead me to reaffirm it. It is possible but not necessary, therefore, that wavering intention may lead to the setting aside of a decision. Uncertainty about

what one will do does not necessarily imply uncertainty about what to do.

The second way in which I believe that those I am criticising are substantially wrong about decision making concerns the relationship between deliberation and decision. To characterise the making of a decision as a passage either to a state of certainty about what one will do or to a state of intending to do it implies that deciding is not a discrete moment - a singular act - but merely a relation between states: a relation of transition. If this were so, however, we would need to know what it is that brings the transition about: what it is that causes deliberation to cease and intention to commence. If ***decision*** is no more than a way of describing the transition from deliberation to intention, it cannot be said to be the cause of the transition. The only alternative is that there is something in the state of deliberating that brings it to a conclusion. But what could this be? That is to say, what is there in reasoning between alternatives about what to do, which could cause the process to self-terminate? The only possibility, it seems to me, is that as a result of careful deliberation a moment arrives at which an intention is formed because there is no longer any doubt about what to do. This is O'Shaughnessy's view. He argues for it by asking us to imagine a situation in which a jury of one has to decide whether or not to bring in a guilty verdict. According to O'Shaughnessy, the juror's enterprise of trying to make up his mind about the question of guilt and that of making up his mind about what to do would be identical, but the event that completes the first enterprise will not be that which completes the second:

> the event that completes factual rumination is cognitive crystallisation, the onset of cognitive commitment; whereas the event that completes practical rumination is the onset of practical commitment, which is the onset of an intention state. But no coming-to-believe can, under any description, be a coming-to-intend, even though these ruminative procedures are the one procedure under different descriptions. In short, the incident of 'deciding whether' is necessarily distinct from the incident of 'deciding to do', even though these two milestones are reached along the same road.
>
> And so three distinct events must be crammed into the one instant. Namely: the termination of the theoretical/practical ruminative procedure, which is distinct from, and the terminus of the cause of, the event of cognitive crystallisation; which is in turn distinct from and cause of the event of practical commitment. Note, in conclusion, that deciding does not as such require the occurrence of a preceding process of rumination; for a man can go to bed undecided and wake to a state of

> decision, without there needing to be rumination during sleep. All that may be required is that the mental dust should settle. In any case, that practical uncertainty should give way to practical commitment. Deciding is the resolving of doubt.[28]

As I said, O'Shaughnessy is careful to stress that intending and coming to intend (deciding) are distinct from believing and coming to believe. But what is more significant is what he claims that coming to intend, deciding and coming to believe have in common. Coming to believe - 'the event of cognitive crystallisation' - is caused by the termination of deliberation, and coming to intend - 'the event of practical commitment' - is caused by the event of cognitive crystallisation. Presumably when ***deciding to*** encompasses more than ***deciding that,*** practical commitment will not follow solely from cognitive commitment, but from that taken together with practical deliberation, or as we should say, it would follow from practical deliberation, of which one element would be cognitive commitment. What should be clear, at any rate, is that O'Shaughnessy sees practical commitment as something like a moment of realisation: a moment in which deliberation is no longer necessary. O'Shaughnessy's choice of 'cognitive crystallisation' to describe the onset of cognitive commitment is very apt. Crystallisation suggests the culmination of a process that has been going on during deliberation, while at the same time implying the beginning of something that will endure thereafter. Its existence is not discrete but describable only in relation to the process of deliberation and its replacement by belief. Likewise with the onset of practical commitment, in O'Shaughnessy's view there comes a moment when intention crystallises out of deliberation - even with an intervening period of sleep - or when the mental dust settles: a moment in which it is clear what to do.

That O'Shaughnessy does view decision as being like realisation is confirmed by his claim that ***deciding to,*** just like ***deciding that,*** is not an activity.[29] Of course ***deciding to*** is not an activity just in the trivial sense that it is a singular act rather than a process continued over time. But what O'Shaughnessy means is that ***deciding to*** is not an act, which is to say, it is not something we do but rather something that happens to us as a result of what we do, like realisation.

It is sometimes true that practical deliberation unambiguously reveals what we should do, and that our desires are so unambiguously in accord with the result that intention to act is formed seamlessly with that result. It is sometimes true, but often not. Often enough deliberation has no obvious terminus. I may deliberate for a time and reach a tentative conclusion about what to do, but then go on to consider new alternatives, or

new reasons for other alternatives; or I may review whether the tentative conclusion was properly arrived at. Deliberation is sometimes of a character that will allow it to be carried on indefinitely. Deliberation that could otherwise be completed within a certain time often has to be concluded more quickly because of time constraints. Where a natural terminus for deliberation has not been reached, and where no intention to act has sprung into life unbidden, a definite act of resolution must be taken to bring deliberation to a close.[30] Decision is the act of resolution.

Dennett raises doubts about the notion that deciding involves a definite act or moment:

> 'I have decided to take the job,' one says. And very clearly one takes oneself to be reporting on something one has done recently, but reminiscence shows only that yesterday one was undecided, and today one is no longer undecided; at some moment in the interval the decision ***must have happened,*** without fanfare. Where did it happen?[31]

Dennett suggests that there is sometimes no definite moment of decision; in which case, presumably, one simply becomes aware or one notices, at some point, that one has decided, or that indecision has been left behind. But noticing that one has decided can only mean noticing that one ***intends,*** for example, to take the job. Suppose then that you are strongly inclined to take a job, but that you are reluctant to do so because it is being offered with 'a nod and a wink', on the basis of patronage, perhaps, and without regard to equality of opportunity or fair competition. Say, then, that you wake up in the morning, and you are aware that you intend to take the job, but you have no awareness of a prior act of decision. You might have the intention because you have temporarily forgotten why you were undecided about it, or because you have forgotten the force of your qualms about it. In that case, it seems to me, in order to count yourself as having decided, you would definitely have to *affirm* the intention (thus overriding your qualms) or abandon it; and such an affirmation would be a definite act of decision. Or, alternatively, you might form an intention to do something you were previously undecided about, and go on to act on it, having forgotten your previous indecision and the reasons for it. In that case you would not have ***decided*** to act as you did.

Strictly speaking, then, when deliberation effortlessly terminates in intention, we ought not to speak of a decision as having been made. But since the same result - determining what to do - is achieved either by deliberation coming to a natural resolution or by its being actively brought to a resolution, this looser usage is intelligible. And it remains the case

that even with decision making in our wider sense, wavering intention does not necessarily imply that there should be a return to deliberation and therefore does not necessitate abandonment of the 'decision'.

According to Ginet, and Hampshire and Hart, as we have seen, it is not possible to know what one will decide while one is still deliberating, because deliberation involves uncertainty, which would be terminated by knowledge of what one will do. Taylor also argues that foreknowledge of what one will do is incompatible with deliberation since deliberation implies that one does not know what one will do. I have argued that the making of a decision is not properly characterised as a transition from uncertainty about what one will do, and that it is an act rather than a cognitive state. I have also argued, however, that while making a decision does not yield certainty about what one will do, it does involve the resolution of uncertainty about what to do. Isn't it impossible, then, to be uncertain about what to do while knowing what one will decide?

If I were to make a prediction about how I am going to decide, or to be told of such a prediction, then either I could not believe the prediction or I could no longer deliberate about what to do. But since deciding is the act of resolving deliberation by marking an intention to act, if this were true no prediction of a decision could be true. This is difficult to accept. Decisions are surely not all so erratic and unforeseeable that one cannot have a good idea in any instance about what decision one will make. One might have a good idea, but not knowledge, according to Ginet: 'the concept of a decision does not allow the possibility of a person's knowing what his decision will be before he makes it'.[32] According to Taylor it is perfectly possible to predict what one will do and to know what one will do thereby, but in that case one is past ***deciding*** what to do.[33]

Suppose, however, that a person is faced with a decision he has had to make several times before: if he made the same decision on all previous occasions and has no reason to believe that there has been any change that might significantly affect how he will decide this time, could he not know how he will decide? In that case would we say that in calculating that there has not been any significant change that would affect his decision, he has deliberated and decided rather than predicted what he will decide?

Pears offers a counterexample to this suggestion, in which a chess player who is faced by the same simple position as a friend of similar skill in a previous game predicts that he will make the same move.[34] Can the player ***predict*** that he will make that move without at the same time ***deciding*** that

he will do so? If the prediction is based solely on past experience of his own decisions and on what move his friend has made, he will not know why his friend has made the move and consequently why it is the move that he would also make. It is compatible with the prediction, therefore, that he works out how the move is arrived at before he makes it. Pears points out that since the practical problem about what move to make contains a theoretical problem about what is the optimum move given the position, then on the basis of previous experience there can be a high degree of inductive certainty about the prediction. Pears accepts that it may be wrong to describe the process of working out as one of deliberation, or what follows from it as a decision, given the lack of uncertainty, but argues that in discovering how the move fits the position and the player's desire to achieve the swiftest checkmate, 'when he feels the direction of his desire, he is making the decision his own'.[35] In that case, Pears argues, even without the uncertainty that goes with deliberation, there is 'more that is the same' between the chess player's problem solving and what we would strictly allow to be deliberation, and between his making the move and what we would accept as his deciding to make the move.

Pears argues that what raises the chess player's workings-out above the level of a merely theoretical exercise is the connection between the workings-out and his desire for the swiftest checkmate. But if the player is completely certain about what move to take (and if he is certain what move to take he is ***eo ipso*** certain that this is the move that will best realise his desire), then his working out why the move is the right one ***is*** a mere theoretical exercise, which he cannot regard as helping to realise his desire in any way. It would be as relevant to his desire to win as would adding up the number of pieces that are left on the board or working out the ratio of remaining pawns to higher pieces. Working out the move certainly cannot serve to ***confirm*** it, because if we are certain of something we do not require confirmation of it. The only way that the player can view his workings out as in any way connected with his desire is if he believes those workings out could help to realise his desire; and he can only believe that if he is, in some degree, less than certain that it is the right move. Since he desires the swiftest possible checkmate and is committed to making the move that is most likely to realise that desire, if he is in any way uncertain about what the best move is, he cannot regard the move he will make as completely certain. If he cannot be certain about his prediction, is it true that he is thereby prevented from knowing it? As Pears has argued, there can be a high degree of inductive certainty about a prediction such as this. If there is genuine deliberation about the move there cannot

be total certainty, but wouldn't a high degree of certainty about what the move will be, although short of complete certainty, be enough to say that the chess player knows what the move will be?

A sceptic might argue that for something to count as a genuine object of knowledge requires that it be completely certain. But this would fix an artificially tight requirement for what is to count as knowledge, which, arguably, even analytical truths would not satisfy. The sceptical argument would certainly rule out the possibility of knowledge based on induction. I take it that those who have argued that it is not possible to know what one will decide before one does decide imply thereby that it *is* possible to know other things by prediction: if not, the claimed contrast between decisions and other events would be vacuous.

Surely, however, even if there is good inductive evidence for a prediction, which we would usually accept as conferring a high degree of certainty, that certainty would still be undermined if there were other reasons for doubting the prediction. And if we were deliberating about what decision to make, wouldn't that be sufficient reason to doubt the prediction, given that deliberation involves uncertainty? The answer is no, for the reason that current deliberation and uncertainty can be judged against past deliberation and uncertainty. If the evidence is that the chess player has always or nearly always decided in the same way as his friend, no matter how uncertain he felt and how much he deliberated, then current deliberation and uncertainty give him no reason to doubt the prediction.

It might be suggested that if a predicted decision fails to occur, then even if there had been a high degree of inductive certainty for the prediction it would still have been false and the chess player could not have been said to know it. Therefore, inasmuch as the chess player is uncertain about his decision, he entertains the possibility that what he has predicted might not happen, and therefore that it is not something he could know. Again, the answer to this is that we would not ordinarily say that we do not know that something is the case, or will be the case, just because there is some small uncertainty about it.

The uncertainty of deliberation would not give reason to doubt a prediction based on inductive evidence about the inefficacy of previous uncertainty and deliberation. What it would do, as I have already conceded, would be to impinge on the player's certainty about the prediction. But by how much? If the player were hardly certain of the prediction of his decision, then we would not say that he knew it to be true. But why should the uncertainty involved in deliberation necessarily reduce the player's conviction about the prediction enough for us to say that he would not know the prediction? Substantial uncertainty when deliberating implies

substantial uncertainty about the prediction. But why should we think that deliberation must involve substantial uncertainty? I see no contradiction in deliberating about what to do while being ***fairly sure*** what those deliberations will produce. Someone, for example, who has thought about what to do in a particular situation and has returned again and again to the same unpalatable conclusion might, nevertheless, continue to deliberate in the slight hope that there is something he has previously missed. There is nothing contradictory in the supposition that the chess player deliberates with only a low degree of uncertainty (thus with a high degree of certainty) about the outcome, and therefore that he can know, without being completely certain, what he will decide before he decides.

Another objection would be that however certain the chess player is of what he will decide, to that extent he ***has*** decided. Likewise, that in whatever degree he remains uncertain, to that extent he remains undecided. There are, indeed, a series of English locutions that do seem to express the idea that one is not undecided, but not wholly decided either. They include such expressions as 'I've almost made up my mind to', 'I've half a mind to' and 'I'm almost decided'. Such phrases, however, are misleading. What they really refer to are not, in the first place, decisions and intentions, but reasons and beliefs. To say that 'I'm almost decided' is to imply that there are good reasons for deciding to do something, on the basis of which I expect that I will decide to do it, although I have not yet decided, which is to say that I have not yet resolved to do it. If I describe myself as having 'half a mind to' do something, I would be understood by that again to mean that I have good reasons for doing it and, perhaps, that I desire to do it, but that there are also reasons for not doing it, in consequence of which I might not do it. It is not possible to half-decide to do something: one is either decided or one is not. Deliberation is either going on or it is not. Where there is deliberation, decision has yet to come (including when a previous decision has been set aside), and where a decision has been made, deliberation has been concluded.

As a counterexample, suppose I decide to get up at 7 a.m., but then I remember that whenever I have previously decided to do this I chickened out.[36] Suppose in addition that by thinking about my past failures of resolve I am led to reason about what it was that led me to abandon getting up at 7 a.m.: that the room was cold, that I was still tired and that it would have been an unpleasant experience to get up that time. In that case I might be led to wonder whether I shall carry out my decision this time. I might focus on my reasons for deciding to get up at 7 a.m. (that I would get a lot more done, for example) in an attempt to re-establish my shaken sense of resolve. If the upshot of all this is that despite a strong desire that I should get up at 7 a.m., and a resolve to be firm with myself when the

time comes, I am just not sure what I will do, couldn't it be claimed that in that case I would be 'half-decided'? After all, I will not have revoked my decision: in a sense it would still stand. But if I am really not sure whether I will carry it out, I cannot honestly claim that I fully intend to carry it out, and in that case can I regard myself as 'fully decided'?

The answer is that if deciding is just the passing from uncertainty to certainty about what one will do, or a passing from deliberation and uncertainty to intention, then if one does not fully intend, one will not be fully decided. As I argued above, however, wavering intention does not necessitate indecision. I might come to doubt whether I will carry out a decision, through laziness or faintheartedness or whatever, but if I made the decision with the intention of overcoming laziness and faintheartedness, and if I continue to regard them as tendencies I ought to resist, then there may be no reason for me to return to deliberating about what to do. In that case I may waver in my intention, while remaining firm about my decision: I will know what to do without knowing what I will do.

It is true, however, that if I have not yet decided to get up at 7 a.m., and reflecting on past evidence I am doubtful about whether I would get up then even if I did decide, for as long as I am doubtful I will be prevented from deciding to get up at 7 a.m. The reason for this is that deliberation can only result in a decision by marking an intention, and it is a conceptual impossibility that one does this while being aware that one has doubts or beliefs that prevent one from having such an intention. If one were thus prevented, one might conclude that getting up at 7 a.m. is the right thing to do, and what one wishes one would do, while lacking any confidence that it is what one will do, or perhaps being confident that it is not what one will do. In that case deliberation would have been resolved by practical judgement rather than by a decision.

I conclude therefore that while it is impossible to be completely certain about what one will decide before one does decide, it is possible, in certain circumstances, to know with a high degree of certainty what one will decide before one does so. There must be substantial uncertainty about a prediction before we should say that we are prevented from knowing it to be true. And neither can it be said that however uncertain a person is, to that degree they are undecided: one is either decided or one is not.

DIFFICULT DECISIONS

If it is possible to have causal knowledge about what one will decide before one does decide, providing there is a high degree of certainty about the prediction and a correspondingly low degree of uncertainty in

deliberating, it must also be possible, according to the analysis set out earlier, to experience such decisions as caused. If a prediction is believed to be valid it can be the object of a causal expectation. Therefore, even if the virtuous official's refusal of bribes is to be rejected as a putative example of a decision being experienced as caused, on the grounds that it is not preceded by real deliberation and uncertainty, the same objection cannot be made in respect of cases such as that of the chess player where there is real deliberation but with a low degree of uncertainty about the outcome. It is possible, therefore, if I am deliberating with a low degree of uncertainty about what I will decide, that I can experience my decision as an effect. But where there is significant uncertainty, causal expectation is ruled out. Still, where there is little uncertainty about what a decision will be, can the decision be one that involves the experience of freedom in the sense that incompatibilists take to be threatened by determinism? This is the sense, as I have said, of it being ***ultimately*** up to us how we choose and decide.

The experience of freedom to decide attaches, as I mentioned above, to deliberation which involves real struggle and uncertainty in arriving at a decision. The sense of it being truly up to me how I decide hinges, in fact, just on that sense that the decision could go either way and that there is nothing and no one other than myself that can decide it. If experiencing something as caused requires that one have a causal expectation of it, and since causal expectation is incompatible with a significant degree of uncertainty, then it is not possible both that a decision be experienced as caused and that it be experienced as ultimately free. A difficult decision, which is to say, a decision that is preceded by significant uncertainty, cannot be experienced as caused.

Our inability to experience a difficult decision as caused, however, is insufficient to clinch the incompatibilist argument. The argument is that the truth of determinism would render our experiences of decisions, especially difficult decisions, illusory. As I argued above, if our experiences of difficult decisions are to be thought of as illusory, on account of their being caused, there must be some logically possible veridical experience corresponding to what we experience as difficult decisions.[37] What would such experiences be? Since what it is about the experience of making a difficult decision that conflicts with our experiencing it as caused is the substantial uncertainty of the deliberations that precede it, a veridical experience of a caused 'decision' would be one that lacks this uncertainty, or reduces it to the minimal level we have found to be compatible with causal expectation and foreknowledge.

On the face of it this is not so difficult to imagine: indeed we have

already done so. To experience difficult decisions as caused would be to experience them as ones in which uncertainty is reduced to the level of that experienced by the virtuous official or the chess player. But in that case, of course, they would no longer be difficult decisions. What characterises difficult decisions is uncertainty about what to do, and this is not just a matter of how they are experienced but of how they are made. Uncertainty is not just a feeling accompanying deliberation but something that enters into it and affects its progress. There is therefore a sense in which, if our experiences of difficult decisions as free are illusory, it would be impossible for us to have veridical experiences of them. This, however, would not undermine the claim of illusoriness anymore than trembling in one's seat as a big-screen Dracula rises from the grave would undermine one's belief that there are no such things as vampires.

Veridical experiences involving causal foreknowledge of what we now experience as difficult decisions would presumably be ones in which there would be causal expectation involving a high degree of certainty about what the decision would be. Uncertainty would be abolished or significantly reduced by a clear appreciation of the situation, one's desires, values, goals and so forth, together with an understanding of the laws linking them to particular kinds of behaviour, or at least by some kind of relevant and reliable inductive evidence. In that case causal expectation and 'decision making' would tend to merge. The moment of decision would largely coincide with the causal expectation of it, or expectation would immediately yield to decision. There could be a separate moment of decision just to the extent that deliberation is directed on some small uncertainty about what to do. Again, if causal expectation were to yield immediately to decision, this would give us not just a changed experience of decision making, but a change in what making a decision involves. The important point, however, is that the putatively illusory experience of uncertainty and of it being inescapably up to me how to decide would have been dispelled.

But suppose I did understand all this, and did, for a time, experience my decisions as caused: wouldn't it still be possible for me to resist an expected decision? Couldn't I defy inductive certainty and do something else? Incompatibilists would say that the belief that we could do this would be a reversion to belief in one's freedom to decide, revealing how entrenched it is in our experiences, beliefs and attitudes. But they would have to say that if determinism is true, the belief that I could decide differently is false. If such a state of affairs were to obtain, I would supposedly find that attempts to act on my belief that I can behave contrary to causal expectations would fail. Thus, through a clear appreciation of the

circumstances, knowledge of my character, the laws governing my behaviour, etc., I might form a causal expectation with a high degree of certainty that I will say yes to a glass of port at the end of a meal. If it should seem to me that I might defy this causal expectation and just settle for coffee, I would be very likely to find out, if I tried to do so, that I could not.

But this seems absurd. Is the reason for that, as incompatibilists would say (as well as those, like Ted Honderich and Galen Strawson, who argue that we have indeterminist attitudes), just that indeterminism is so entrenched in how we think that it is unimaginable or very difficult for us to throw it off? Or is there something wrong with the incompatibilist argument that has led to it?

HOW MUCH CAN WE KNOW TO BE TRUE?

Supposing I do have a causal expectation that at the end of the meal I will 'decide' to have port. My expectation is grounded in dependable inductive evidence about my post-prandial choices on a number of other occasions. But supposing I form a desire to thwart the expectation: what then? Perhaps I had similar desires on previous occasions but still went on to decide in the way I had expected. In that case my expectation would still have good inductive grounds. Suppose, then, that I form a desire that for once my desire to thwart the expectation should win out.[38] I might never have had such a desire before, and so might not have any evidence on which to base a prediction; and even if I had, might I not have an additional desire about that? I could conceivably end up with more desires than I could make sense of, and enough, certainly, to prevent my clearly identifying my mental state as being of a particular type that would allow me to form a causal expectation about what decision I will make. But even were I not to form a multiplicity of desires-about-desires-about-desires, whatever prediction I make about my decision, my knowing the prediction could, for all I know, affect how I decide; and this is not something I would have taken into account in forming the causal expectation. Nor, of course, can the problem be gotten around by taking into account the effect of knowledge of the prediction on my deliberations, because the effects of knowledge of the revised calculations must also be taken into account, and so on, *ad infinitum*. This is not simply a practical difficulty but a logical limitation arising from the fact that beliefs about one's deliberations (or other mental states), and whatever expectations or predictions one draws from them, can in principle affect those deliberations.

This logical obstruction to complete knowledge of one's deliberations,

and any prediction of a decision or action that is inductively inferred from them, has been widely discussed.[39] According to Oldenquist, to know that a given set of conditions are causally sufficient for a decision, one must also know that this knowledge is not itself part of a set of causally sufficient conditions for the decision, but to know that one would also have to know that one's knowledge of it is not also part of a set of causally sufficient conditions for the decision. Infinite regress clearly follows.[40] Is it, then, a conceptual impossibility that one might have causal knowledge of a decision before one makes it?

It does seem possible that one can sometimes know with less than absolute certainty that one's deliberations will not be affected by one's knowing what they are, or by a prediction about their outcome. In the case of the chess player, for example, there would be no reason to expect that knowledge of the predicted decision might affect the decision; nor would we expect this in the case where an agent deliberates about what to do in the small hope that there might be an alternative to the unpalatable conclusion of previous deliberations. One could not be completely certain that knowledge of a prediction or a causal expectation would have no effect in such cases, but as I have argued already, complete certainty is not required for knowledge or for causal expectation.

Where logical indeterminacy about a decision would become much more significant is in cases of conflicting desires, values and goals, of the sort one experiences in making difficult decisions. If one is trying to decide what to do and there are two or more strong and evenly matched sets of contending considerations, then whatever it is that causes one set to prevail (whether that be an additional consideration, or some particular feature of one of the sets of considerations, or that one set of considerations weighs a little more heavily than the other, or that the need to make a decision introduces a degree of capriciousness into the making of it) will in that case be understood as the singular cause of the decision. If the opposed considerations ***are*** fairly evenly matched - which they must be in order for there to be uncertainty and struggle - then whatever tips the balance between them, so to speak, is going to be a fairly small consideration by comparison. Any agent who became aware of this would also be made aware that the effective cause of her expected or predicted decision would be slight or trivial. It would be difficult to know with any degree of certainty that such awareness might not significantly affect one's mental state sufficiently to thwart the expectation.

It is still conceivable that an agent who does form a causal expectation about what she will decide also forms a resolute attitude that the expectation will not be allowed to affect her assessment of reasons for deciding:

that she will, so to speak, mentally bracket off the causal expectation from her reasonings about what to do. But she could not accomplish this in respect of knowledge of the kind of prediction-licensing mental states we have been considering. According to the incompatibilist account I set out above of what veridical experiences of decision making would be like in a determined world, the uncertainty of deliberation would be replaced by assessment of desires, values, goals, and their various weightings, together with the circumstances and the relevant causal laws, yielding a causal expectation about what will be decided. However, since reasoning about what to do is not just instrumentally responsive to desires, values and goals, and their weightings, but can evaluate and compare them in such a way that their respective weightings can be altered, one cannot know that knowledge of one's desires, values and goals will not significantly affect them. If I were to discover, for example, that my desire for the approval of others weighs more heavily with me than my sense of duty, there is no way I could be sure that this information, or any feelings and reasonings occasioned by it, would not alter the balance between them. There is no way, in other words, that one could know for sure, where there are conflicts in motives, that those conflicts can be causally insulated from one's knowledge of them.

For this reason, where there are conflicts of motives of the kind that make decisions difficult, any identification of an agent's deliberative processes as being of a particular causal type cannot be true for her. It is logically impossible, therefore, where there are such conflicts, that an agent should be able to accurately identify her occurrent motives as belonging to particular sets of causally sufficient conditions. It follows that where this is the case it is impossible for an agent to reliably identify a singular cause from among her deliberative processes, or to form an accurate causal expectation on the basis of it, since she could have no confidence that the requisite background conditions are in place.[41] If no accurate causal expectation can be formed where one's motives give conflicting indications about how to decide, then a veridical experience of a difficult decision as caused is a logical impossibility.

An agent could form a causal expectation about what decision she would make, even where she has conflicting motives, but such an expectation would be either untrue or invalid: as I indicated earlier, it is possible to experience a causal relation between events where none exists. Since no identification of the agent's occurrent deliberative processes as being of a particular causal type could be true for her, either her expectation would be falsely grounded in a belief that her deliberative processes are of a particular causal type, in which case it might be *validly* derived from a

causal law or inductive evidence connecting that causal type to the expected decision, or, if her expectation was not derived from a true causal law or inductive evidence, it might be coincidentally true but would not be valid. In neither case would an experience of a decision that involved such a causal expectation be veridical.[42]

THE FUTURE

It might be thought that even if we cannot have causal knowledge of difficult decisions we have yet to make, and even if our experiences of making such decisions are veridical, the burdensome thought remains that if determinism is true we must accept, as we struggle over difficult decisions, that their outcomes, even though we are prevented from knowing what they will be, are already settled. Our futures would be fixed.[43] But to say this is to interpret the logical limitation on knowledge of future decisions, set out above, as merely closing off access to a set of truths about them. For an agent trying to make a difficult decision, or contemplating doing so, there are no such truths to have access to.

MacKay has argued that it is logically possible for an observer who is fully informed about an agent's brain processes, together with any environmental stimuli and relevant law statements linking them to particular kinds of behaviour, to predict the agent's decisions. He also argues, however, that such predictions, although true for the observer (providing he does not reveal the prediction to the agent), will be logically incredible for the agent. This is because knowledge of a prediction or of the deliberative processes on which the prediction is based (or their neural correlates) can causally effect those processes and their outcome. Any prediction an observer might make, therefore, can only be true if it is not revealed to the agent. MacKay concludes that no prediction of an agent's choices or decisions can be true for her.[44]

I have argued that where there is little uncertainty about a decision, it does seem possible that one could know in advance what it will be. But where making a decision involves real struggle and uncertainty this is ruled out. It is not merely that we cannot know the outcomes of difficult decisions before we make them, therefore, but that no antecedent descriptions of such decisions can be true for us. Our futures, inasmuch as they are affected by our deliberations and difficult decisions, are not fixed, but open.

It might be argued that although it is not possible to know the outcome of a difficult decision before it is made, it is logically possible that an

agent be subsequently shown that she could not have decided differently. It is also possible that an agent could know before she makes a decision that this information could be revealed to her after she has made it. In that case she might validly infer that although she cannot have antecedent causal knowledge of her decision, she can know that only one decision is possible for her, although she does not yet know which. But since the statement that a particular decision is impossible (even if one doesn't know which decision) is inferred from descriptions of causal processes that cannot be true for the agent who is deliberating about whether to make that decision, then such a statement can only be validly and truly asserted by someone other than the agent. Or, rather, it ***can*** be validly asserted by the agent, but only in a manner of speaking in which the logical status of the statement is such that she disregards her own attempts to come to a decision. When she does this she supposes, ***sub specie aetemitatis,*** as it were, that there are laws governing her behaviour and environment that would allow a valid and true prediction of whatever she might do. One logically adopts a different relational position when one is considering what to do. In that mode, or from that perspective, statements that assert or deny the impossibility or necessity of particular difficult decisions, and any statements that can be inferred from them, cannot be truly and validly asserted, since to know them as valid and true would depend on antecedent causal knowledge of the decisions, and such knowledge could never, for reasons that are now established, be possible for the person whose decisions are predicted.

Bivalence fails, therefore, in respect of statements about difficult decisions one has yet to make. Whether a statement about a future difficult decision has a determinate truth value is relative to the person who might assert or entertain it and dependent on when it is asserted or entertained. Those who find themselves reluctant to accept this, and who are still inclined to think that, even though one cannot know the outcome of a difficult decision before one makes it, one can, if determinism is true, still have antecedent knowledge that the outcome is already settled, should consider the following argument. The claim that determinism entails, in respect of any difficult decision, that there is only one way an agent can really decide implies that an agent struggling to make a decision could truly assert 'although I cannot know how I will decide before I do so, I know that there is only one way I can decide'. But no agent could truly assert this. Suppose that you must decide between ***P*** or ***not-P.*** If your decision is already settled you could truly say 'It is already settled which way I will decide, although I do not know how I will decide.' To say that it is already settled either that ***P*** or that ***not-P*** implies that there exists a

state of affairs which is referred to by *'P* will happen', or if not by *'P* will happen', then by *'not-P* will happen' or *'P* will not happen'. To say that you can know that such a state of affairs exists entails that you can know that nothing you do will be inconsistent with its existing.[45] But if you cannot know how you will decide before you do decide, then you cannot know that you will decide for *P* and you cannot know that you will decide for *not-P*. Since deciding for *P* is inconsistent with *'P* will not happen' and deciding for *not-P* is inconsistent with *'P* will happen', and you cannot know which you will decide for, you cannot know that you will not decide in a way that is inconsistent with *'P* will not happen' or that you will not decide in a way that is inconsistent with *'P* will happen'. You cannot know, therefore, that there exists a state of affairs that is referred to by *'P* will happen', or if not by *'P* will happen', then by *'not-P* will happen'. Since you cannot know that such a state of affairs exists, you cannot know that it is already settled how you will decide. In respect of difficult decisions an agent has yet to make, the distinction between not being able to know something and there being nothing to know collapses.

Although it is possible, in a manner of speaking, to adopt a disengaged perspective on one's actions and decisions from which it is possible to assert that the outcome of one's deliberations is already unalterably settled, it is important to remember that this perspective is no more than *a manner of speaking* (or thinking). Planning, acting and deciding, by contrast, are inescapable for us; or at any rate they are a good deal less escapable than taking the disengaged perspective in which one imagines that one is observing oneself. From our perspective as deliberators and deciders, therefore, it is literally possible for us to decide either way. Inasmuch as one's future does depend on one's decisions (which is very often), especially one's difficult decisions, it is, so to speak, an open book.

It also follows that since we can, when confronting difficult decisions, decide either way, and that we cannot truthfully regard any existing or antecedent state of affairs as necessitating us to decide either this way or that, we are bound to regard our contributions to such decisions (and to any actions that involve such decisions), *qua* agents or deciders, as ineliminable: our difficult decisions really are 'up to us'.

CONCLUSION

The incompatibilist argument that if determinism is true our experiences of being free to decide are illusory therefore fails. In claiming that our experiences of making difficult decisions are incompatible with determinism,

incompatibilists have tried and failed to give expression to something about the way we experience making difficult decisions. That something is the sense that such decisions are entirely up to us. This sense, however, is not one in which we experience decisions as positively ***uncaused,*** but rather that we do not experience them ***as caused.*** This distinction is important, since if our experiences of decisions as 'up to us' are merely ones in which we do not register them as caused, this would only be incompatible with determinism if there were some logically possible way of experiencing such decisions veridically that would include their being caused. But there are no such possible experiences, and our real experiences of making difficult decisions are a good deal more ***metaphysically innocent*** than incompatibilists have taken them to be. We may add that claims that we all have entrenched indeterminist attitudes in respect of the initiation of actions have also been partly owed to misidentification of our experiences of deciding as being indeterministic in content.

I have shown how an agent might plausibly be described as experiencing her decisions as caused, where such decisions follow effortlessly from stable virtues of character. I have also shown that a person might experience her decisions as caused where there is a negligible degree of uncertainty in the deliberations that lead to the decision. Decisions such as this, however, are experienced as free only in the sense of not involving any obstacle, hindrance, constraint or compulsion, and not in the sense that incompatibilists have thought to be essential to what most of us experience or mean by freedom: the experience of it being inescapably up to us how we decide. This experience does attach to difficult decisions, which is to say, decisions that are arrived at by struggle and uncertainty. Because of the conflict of desires, values and goals that makes such decisions difficult, however, no identification of occurrent motives and deliberations with particular causal types can be true for the agent. In consequence, an agent with conflicting motives cannot form a valid causal expectation based on causal identification of those motives. Veridical experiences of difficult decisions ***as caused*** is a logical impossibility.

Some incompatibilists might be tempted to argue that since it is a logical impossibility that we experience difficult decisions as caused, it is a logical impossibility that we should consistently believe in determinism. In that case they would need to show how inability to experience a difficult decision as caused necessarily implies belief that it is uncaused. But they would then be committed to a series of absurd claims, such as that no one can believe in temperatures of 1000 °C, curved space or echolocation.

It is no consequence of determinism, therefore, that our decisions, and

our futures more generally, are fixed in advance (only in ways that we as deciders can have no cognitive access to). If no description of a difficult decision I have yet to make can be true for me, nor any details of my future that might depend on it, then for me *qua* decider such descriptions, and the future that depends on them, are not fixed: for me there are no (non-tautological) truths about the details of difficult decisions I have yet to make. In addition, since we can never have valid and truthful causal expectations or predictions of difficult decisions we have yet to make, we have no order, sign or evidence to tell us what to do. We are, in Sartre's sense, *compelled* to decide freely, even if determinism is true.

Two problems remain about whether we can truly be said to have *ultimacy* of the kind that incompatibilists have sought to give expression to. The first problem concerns our categorical ability to act or decide differently than we do. I have argued that since there are no truths about difficult decisions and any actions that might involve them, it follows that we really do have a categorical ability to decide either way. But as compatibilists will remind us, if we do not regard our future decisions as fixed by antecedently existing sets of reasons, how can we claim that the decisions 'we make' will be either rational or under our control? The second problem is that although it has been established that our difficult decisions cannot truthfully be regarded as entirely owed to anything or anyone other than ourselves, this is not enough for our decisions to be 'truly up to us', since it remains possible that there is no 'us' (no real agents) for our decisions to be truly 'up to'. Both problems are bound up with difficulties about the nature of actions and agency and it is to these that I now turn.

5 What are Actions?

Is agency an illusion? It seems as if it is me who ***writes*** these words, who ***presses*** the keys, and who ***does*** many other things besides. But after all, according to the view I have assumed in the previous chapters, my actions are just the end results of causal sequences. They may in some important sense be free, but if they are caused by mere events and can be explained as such, then to say that they are caused by me seems an unnecessary addition. In everyday talk I may be warranted in saying that I did this or that, but all this comes to is that certain events have occurred and their occurring has caused certain other events. Does that matter? Has anything been lost? If naturalistic accounts of actions as behaviour that is caused by certain kinds of antecedent events or states seem to erase our contributions as agents, perhaps it appears so only because we fail to keep in mind that such causal sequences are ***constitutive*** of our acting. This has been the response of Hobbes and other naturalist and compatibilist thinkers, who have sought, with varying degrees of sophistication, to show human actions as natural events, to be distinguished from other natural events solely according to what causes them and the media in which they occur.

As critics of naturalism have pointed out, however, such accounts fail to do justice to the ordinary sense a person has of herself as acting and doing. Against the Hobbesian view that one is always caused to act by one's strongest motive, Thomas Reid argued that we each have the power to act without motive, against motive and in favour of weaker motives.[1] Whether or not we really have such a power, Reid's view is certainly faithful to how we understand and experience our own roles as agents in relation to the motives that give us reason to act. This can easily be seen in respect of situations where we have conflicting motives about what to do. According to the Hobbesian account, we are determined to act by the strongest motive, or in other words, the motive chooses itself by main force. As Reid rightly protests, this is not how we experience ourselves as determining what to do when we are ***confronted*** by opposing motives. When that happens, it seems that we ***adjudicate*** between the motives and ***decide*** in favour of one against another. Determination is made by the man, and not by the motive.[2]

According to Thomas Nagel, there is no hope of reconciling this view with any objective or naturalistic view of actions, since the latter will necessarily treat actions as events and subject to causation by earlier events.[3] Any attempt to marry our conception of agency with an objective

description of actions as resulting from events will always run up against the subjective intuition that when I act it is ***me*** that does so: that I am the source of my actions. The subjective experience of 'agent-causation' is seemingly irreconcilable with any objective description of event-causation.

Just as with free will and decision making, it is argued that if it should turn out that our actions are completely explicable in terms of events and therefore subject to natural causation, our everyday sense of ourselves as able to act and choose must be profoundly illusory. And if this is correct, then so much for the argument that the veridicality of our experience of freely making decisions is untouched by determinism, or any more general reconciliation of free will with determinism, since why should it matter that our acts and decisions are free, if they are not, in some fundamental sense, really ***ours?*** In the two remaining chapters I will show that our everyday understanding of actions and agency can be reconciled with natural causation, and, as was the case with choices and decisions, that resistance to this is partly owed to lack of attentiveness to the contents of our experiences of actions and acting.

Ideas we have about ourselves as agents are rooted in various paradigm cases of acting of varying degrees of complexity. At the most simple and straightforward level we have the experience of being able to control our bodies in some ways, but not in others. If I raise my arm, this is something I do, in a sense that is not true when I sneeze. I seem to know directly and without question when I raise my arm that it is me that does so. How do I know that? What marks the movement out for me as an action? It is not, for sure, a matter of convention: that a linguistic community designates certain forms of bodily behaviour as voluntary and others as involuntary.[4] We know this to be so from those forms of behaviour that overlap the boundary. I may blink either deliberately or involuntarily, for example, and I will usually know what sort of blink it is. For all we can tell about it as a bodily event, however, a blink is a blink is a blink; and there seems to be nothing about a blink, ***qua*** blink, that identifies it as an action or an involuntary happening.

Now consider a more complex type of action. You have woken up and are lying in bed. You are no longer sleepy and the room is warm and bright. You would like to get up, and there is no desire you are aware of not to do so. And yet you lie in bed for the next half hour, intermittently thinking to yourself: 'I ought to get up. I wish I would get up.' At last you think to yourself something like this: 'This is no good at all; I really must

get up.' And then you do. When you rise from the bed it seems as if it is you that does so; and, anyway, that it is certainly not the mental uttering of a sentence that causes it. So why didn't you get up before you did? Nothing else would have changed at the moment you got up. It would not have seemed as if you were released from some enchantment. So why didn't you get up when you wanted to?

Finally, suppose you are faced with an acute dilemma. You must choose between what you believe to be your duty and something else you badly want. It seems that you have the capacity to reflect on what to do, to assess your motives for the opposing courses of action and to come to a decision about what to do on the basis of that. This is not just like raising your arm. It involves the idea that you are able to decide either way, that you can assess your own reasons for action and come to a decision about them, and that you can bring yourself to act on the basis of your decision. Is it really like that? or is it instead that your strongest desire causes you to behave as you do? Is the notion that you can somehow stand apart from your motives and come to an independent judgement about them - even if that means judging in favour of a weaker motive - an illusion?

The three cases express paradigmatic examples of acting, each of which exemplifies different ideas and intuitions we have about ourselves as agents. When I think of myself as standing in judgement of competing reasons for action, able to critically assess my own attitudes and motivation, the idea of agency this involves is more complex than what one thinks of as going on when one simply raises one's arm.[5] It would be surprising, however, if what is required for the complex conception of agency we take ourselves to have in relation to conflicting motives for action were to have nothing in common with the sense that simple actions are our own doing.

I will refer to the respective problems of explicating what is involved in these cases of agency, and determining whether they are veridical or illusory, as the problem of action and the problem of fully human agency.[6] The problem of fully human agency will be dealt with in Chapter 6. I begin with the problem of determining what it is about simple actions that distinguishes them, and enables us as agents to distinguish them, from involuntary forms of behaviour.

DEFINING ACTIONS

In our first paradigmatic example of acting the problem of distinguishing actions from involuntary behaviour has two components: one ontological - having to do with what distinguishes actions as a class of events; the other

epistemic - having to do with how we as agents are able to distinguish them. In attempting to settle the question of what distinguishes actions from involuntary behaviour and how we as agents are able to make the distinction, the obvious way to begin is to attempt to define what is meant by ***action.*** Even if defining what is meant by ***action,*** or by ***acting,*** does not explain how we are able to distinguish actions, it is a necessary preliminary, since if the distinction is to be intelligible to us, our ability to make it ought to cohere with our concept of action.

Any attempt to define, for philosophical purposes, what is meant by ***action*** and its cognates faces problems in coming to terms with ordinary usage. ***Action*** and ***agent*** are often used to refer to inanimate objects, processes and events for which there is no corresponding class of involuntary behaviours. Several philosophers have also thought it significant that our ordinary usage of the terms ***voluntary*** and ***involuntary*** signals something 'fishy' about the behaviour they refer to, thus suggesting that the philosophical use of the terms, in which they do not refer exclusively to fishy behaviour, is profligate and that any problems that depend on it are muddled (and potentially worthy of the epithet 'pseudo-problem').[7] There is nothing fishy, however, in the enterprise of attempting to frame a definition of ***actions*** as contrasted with ***involuntary behaviour.*** The strongest ordinary-language argument against philosophical notions of voluntariness and involuntariness was directed at their use in attempting to frame accounts of moral responsibility that overlook the range and subtlety of our ways of qualifying descriptions of behaviour in order to assess its culpability or the blameworthiness of the agent.[8] But while we may not, in assessing culpability, employ a fixed distinction in our descriptions between actions and involuntary behaviour, there are many unremarkable ways we have of talking that do imply such a distinction (descriptions of what we do, as opposed to what our bodies do, or as opposed to what we suffer or undergo). And it is the distinction implied by the latter kind of action talk that raises the philosophical problem of action.

The concept of action we need to define, therefore, is that which refers to what Hobbes describes as 'animal motion' or 'voluntary motion'.[9] Voluntary behaviour in humans is that which in contrast to other kinds of events is attributable to us as whole individuals, rather than to our component parts. We may follow Aristotle in drawing the distinction thus:

> the stick moves the stone and is moved by the hand, which again is moved by the man.[10]

If a man uses a stick to move a stone he can be described as moving both the stick and the stone, but a description of what brings about the

movement of the stone or the stick can be given without reference to him. The description of what brings about the movement of his hand, by contrast, requires his inclusion. A description of what causes the hand to move could of course be given by reference to the man's muscular and nervous system, without referring to the man ***per se*** playing any part. But such a description would set aside what we take to be a significant fact about the movement of the man's hand: that it is an action. This is not true of the movement of the stone or the staff. Both are brought about by an action, but neither are actions. Actions, in the sense that concerns us, therefore, can be defined as that class of movements whose normal descriptions ***necessarily*** include (at least tacit) reference to the individual who brings them about.[11]

The individual who brings an action about need not be a human. We also ascribe actions to animals, and do so according to the same principle. We distinguish a beast's pursuit of prey from the circulation of its blood or the digestion of its food. The animal's heart circulates its blood, its stomach digests its food, but the animal itself pursues its prey.[12] The class of behaviour that is ascribable to the whole individual rather than to one or more of its parts, however, extends beyond the actions of animals. It also includes the behaviour of various other kinds of self-regulating mechanisms, like plants, thermostats and self-sealing tanks.[13] As a whole we distinguish such behaviour by adopting what Daniel Dennett has described as the 'intentional stance' towards the things we ascribe it to. To describe behaviour in terms of the intentional stance is to explain it in terms of intentional or directed states (desires, beliefs, intentions, hopes, fears, etc.) that render it intelligible.[14] The actions of humans and non-human animals are to be distinguished from the behaviour of other 'intentional systems', according to Dennett, by the greater complexity and range of their self-directedness. Dennett does not believe that the distinction can be drawn according to the real presence in animals and humans of intentional states. Intentional states are to be understood instrumentally as explanatorily useful, and therefore as non-fictitious, but not as identifying fundamental causes.[15]

If actions are 'caused' by intentional states only in the sense that intentional states are explanatorily useful stand-ins for more precise and scientific descriptions of the causes of behaviour, then our prereflective understanding of agency would be partially illusory, because in that case there will be no difference ***in kind*** between human and animal actions and the movements of thermostats or trees.[16] Our concept of action, then, is such that descriptions of actions involve ineliminable reference to voluntary behaviour brought about by individual animals (or their imagined

surrogates). We may add to this, in light of the discussion of intentional states, that actions are intentional or purposive, which is to say that they are done for reasons: I scratch my nose to stop it from itching; rabbits thump the ground to signal danger; and a bear eats a fish because it is hungry. It is debatable whether all actions are intentional, and therefore whether intentionality is necessarily a feature of our concept of ***action,*** and an example of acting unintentionally is given later in the chapter. We can say for now that our concept of action is such that actions are typically purposeful.

This, I think, exhausts what can be said about the concept of action in the sense that interests us. Actions are voluntary behaviour, typically purposeful, carried out by individual animals. The definition, as it stands, does not resolve the problem of how we are able to distinguish actions from involuntary behaviour. To say that when I raise my hand its movement is attributable to me, rather than, say, a nervous spasm in my arm, tells me nothing about how I am able to identify the action as something I do and not as a spasm. Likewise, to say that actions are done purposefully or for reasons tells us nothing about what it means to do things for reasons, or about how one knows that what one does is for a reason, or how one knows what the reason is. The definitional analysis of the concept of action, therefore, does not resolve the problem of what distinguishes actions from involuntary bodily behaviour or how we are able to make the distinction. And it does not do so because the relevant ways we have of talking about actions take the ability to make the distinction for granted.

We need, therefore, to go beyond defining the meanings of ***action*** and related terms to an analysis of what actions really are.[17] To do that will require clarification of how we ***understand*** the various elements of such an analysis: what sense they each have for us. This in turn will require that we think about what it is like to act, which is to say, how we experience ourselves as agents and what understanding of our actions, and their relation to us, that experience involves. In what follows I will attempt to identify and give expression to the various elements in our experiential sense of agency in respect of basic actions. As with the veridicality of our beliefs about making decisions, therefore, where meaning and use do not resolve the issue, we must turn to the phenomenology of actions and acting. This will enable us to identify a set of conditions against which the adequacy of any theory of action can be judged. Identification of the elements of our sense of agency will be drawn from an examination of some of the major contemporary analyses of agency.

CAUSATION

According to causal accounts of agency,[18] what principally distinguishes actions from bodily behaviour is that they are caused in some way (the details of which remain a problem) by the agent's beliefs, wants and intentions. One broadly Hobbesian account of action has the following form:

> ***If*** *A* wants φ and believes x-ing is a way to bring about φ and that there is no better way to bring about φ, and *A* has no overriding want, and knows how to *x*, and is able to *x*, ***then*** *A x*'s.[19]

Is it true, as this implies, that whenever I raise my arm deliberately my doing so is caused by wants and beliefs of the appropriate sort? Sometimes I have raised my arm, or wiggled my fingers, in a rather distracted way, without my doing so being noticeably preceded by any want or belief. It may be that I had a small desire to exercise my muscles, or to amuse myself by the motion of my fingers, but that the desire was so unremarkable that it is difficult to remember, or it is difficult to distinguish the memory of it from the movement it caused. If actions are always caused by antecedent attitudes, however, it seems implausible that our ability to discriminate between actions and involuntary bodily happenings derives from such causes, since one's sense that something is an action can be very much sharper than one's experience of its putative cause.

Consider, then, a sequence of actions in which desires and beliefs play a more obvious role. As I hear the chimes of an ice-cream van and rush to get some change, if I reflect on the sequence of mental events that has just occurred, I may concur with the Hobbesians that hearing the chimes of the ice-cream van caused me to think about ice-cream and to form a desire to get an ice-cream, and also that I remembered that I needed money to get one, which I had to take to the ice-cream van and hand over. I may also notice that hearing the chimes, having the desire, and remembering what to do was followed by my getting some change, going to the ice-cream van, handing over the change and so on. But I should also notice that there was something else the Hobbesian story leaves out: that ***I*** responded to the sound of the chimes, and the having of the desire and the remembering, by forming an intention to do what I went on to do, and that ***I acted*** on that intention.[20] The Hobbesian story can of course be enlarged to include my intention within the causal sequence that led to my getting the ice-cream, but what will still be missing from the account, ***from my perspective as the one who did these things,*** is ***just that:*** that *I* did these

things, that I formed the intention and that I acted on it. If the Hobbesian account is right, then the intention was caused by my desire and my remembering what to do to satisfy it, whereas it seems to me that *I* formed the intention in response to the desire and the remembrance.

A familiar response to this kind of criticism is to say that the occurrence of a set of antecedent conditions, which are said to cause our actions, can constitute ***our*** causing those actions. As it stands, however, the response still leaves motives in the position of causing actions; or causing intentions, which in turn cause actions. The reason we cannot feel ourselves to be included within this is that there is ***as such*** nothing (or not enough) about the motives that are said to cause our actions with which we can identify ourselves.[21]

Suppose I raise my arm because I am seized with an irrational fear that I might have lost the ability to do so, and I have an urgent desire, therefore, to make sure that I still can. When I raise my arm it seems that I do the raising, and not the fear or the desire, or any belief about how to allay the fear or satisfy the desire. The sense that it is me that raises my arm seems to be quite independent of the desire that prompted it. The intuition is borne out if we think again about blinking: this time blinking because one has an eye irritation and one believes that one's desire to be rid of it will be helped by blinking. One can have the appropriate attitude about blinking, and be about to blink for that reason, but be caused to blink involuntarily by the irritation. One could know in that case that one had blinked involuntarily, even though one's blinking had been preceded by a desire to blink. Knowing whether a blink is an action or an involuntary happening, therefore, cannot entirely derive from any attitude that causes it.

Causal analyses of action have on the whole tended to concentrate on the ontological problem about actions, rather than the epistemic problem. Causal theorists might argue that the epistemic problem is a secondary matter, to be dealt with once a satisfactory analysis of action-causation has been arrived at. If it should transpire that our ability as agents to distinguish our actions from our involuntary behaviour is owed to some other feature of actions than the way they are caused, or is arrived at somehow indirectly, such that it is possible in certain situations to misidentify actions as involuntary, and ***vice versa,*** then we might say that the means by which we distinguish our behaviour as active is usually adequate to the purpose but not entirely veridical. To say this, however, would be in effect to abandon a key element in the enterprise of arriving at a satisfactory account of action. If it were to turn out that the sense of agency that informs our ability to distinguish between actions and involuntary behaviour

is only approximate, rough-and-ready, or in some other way less than veridical, then actions would turn out to be only accidentally rational, and our understanding of actions and of ourselves as agents would be in some degree illusory. An analysis of actions in which the solution to the epistemic problem departs substantially from that of the ontological problem must necessarily ***revise*** our sense of agency.

It may be, on the other hand, that advocates of the causal theory have believed that an agent cannot fail to recognise her voluntary behaviour for what it is, just so long as it is caused in the appropriate way by her attitudes. This implies that an agent's recognition of her behaviour as acting is ***inferred*** from her antecedent intentional states. The suggestion conflicts with the sense of ***immediate,*** and thus non-inferential, involvement we have in respect of our actions. Consider, also, the following possibility. A bug flies close to my left eye, but far enough away to leave one who knows about the effects on eyes of close flying bugs to doubt whether it will cause me to blink. If I had wanted to blink just as the bug was getting close, and I did blink, the blinking could have been caused either by an involuntary protection mechanism or by my wanting to blink. Those who believe that recognition of one's actions is inferred from antecedent attitudes will be bound to conclude that in this situation I would not know whether my blinking was an action or not. And having no recollection of ever experiencing the conjunction of a desire to blink and a fairly close flying bug, I do not know whether they would be right about that. My feeling is that were this to happen I ***would*** know whether I had blinked voluntarily, but perhaps the intuition stems from my usually not having reason to doubt whether my behaviour is caused by attitudes.

A further difficulty with the causal account is that it seems possible that someone may want something, or want to do something, know that it is within his power, have no obvious overriding wants, and yet still fail to do it. This might happen because the agent has failed to think through what is entailed by his wants and beliefs. But it is also conceivable that someone has certain attitudes that recommend a course of action he knows to be within his power, that he has judged this correctly, and is aware of having done so, but he still fails to do it. Causal theorists will deny that this is possible. But they cannot deny that it is conceivable, and for as long as we are unable to identify what it is that does the causing, we merely ***assume*** that actions are caused by antecedent attitudes. Any type of state or event that cannot be identified as invariably being followed by an action cannot, without further addition, provide the basis for an adequate causal analysis of actions.

Finally, even if a person's behaviour is caused by a set of beliefs and

desires, which would also be enough to rationalise it as an action, it may still fail to qualify as an action if it is the product of 'wayward causal chains'. A would-be assassin takes aim and fires at his intended victim, but misses. The sound of gunfire, however, causes a herd of wild pigs to stampede the intended victim to death.[22] The death of the victim is caused, we may say, by the assassin's intention to kill; but we would not describe the victim's death as an intentional act of the assassin. Cases of causal deviancy such as this may be overcome by including within a full causal analysis of action, a careful specification of the relationship between the outcome of the agent's basic action and the beliefs and intentions that cause him to act. Among other things, the trajectory between the agent's basic action and its outcome would be required to correspond broadly to how it was that the agent reasoned, believed and intended that his action would produce its outcome.[23]

The problem of 'internal' or 'basic' deviancy, however, has proved to be more recalcitrant. In Davidson's seminal example, a climber wants to rid himself of the danger of holding the weight of another man on a rope and knows that loosening his hold on the rope will rid him of the danger.[24] If the climber is so nervous about the idea of letting go of the rope that he loses his grip, his letting go of the rope will have been caused by his desire to survive and his belief about how this could be secured, but it will not count as an intentional action because it has not been caused ***in the right way.*** But what is the right way? Davidson has said he despairs of being able to answer this question, and to my mind no satisfactory solution to the problem has yet been given.

The possibility of internally deviant causal chains, moreover, tilts the scales against the suggestion that an agent's recognition of her actions is inferred from her antecedent attitudes. It might be suggested that the climber would know that what he did was not an action, because he would have been aware at the same time of the unnerving effect of those attitudes. It is quite conceivable, however, that he could have been unnerved and still have loosened the rope voluntarily and intentionally.

GUIDED BEHAVIOUR

According to Harry Frankfurt, the problems encountered by causal analyses of action stem from a fundamental mistake of directing attention away from the intrinsic features of an action towards the discrete antecedent state or event that is said to cause it.[25] Frankfurt argues that our conception of actions is not such that they are necessarily caused by anything; that

they can be caused by diverse states or events. What distinguishes actions from involuntary movements, he thinks, is the relations the agent bears to those actions while they are occurring:

> during the time a person is performing an action he is necessarily in touch with the movements of his body in a certain way, whereas he is necessarily not in touch with them in that way when movements of his body are occurring without his making them.[26]

The agent is in touch with his actions, according to Frankfurt, if his behaviour is purposive, and this purposiveness is attributable to the agent. Behaviour is purposive when it is

> subject to adjustments which compensate for the effects of forces which would otherwise interfere with the course of the behaviour, and when the occurrence of these adjustments is not explainable by what explains the state of affairs that elicits them.[27]

Thus, behaviour is not purposive because of the causes that initiate it, but because of the guiding causal mechanisms that ensure that it does not deviate from its course. Frankfurt attributes our sense of agency to how it feels to be 'somehow in touch with mechanisms of this kind'.[28] His account, therefore, treats the ontological and epistemic problems as bound up with each other.

It is no doubt important to our sense of agency that when we engage in sustained actions, we experience them, or have a sense of them, as continuing under our control. There is good reason to think, therefore, that such guiding mechanisms are important to our sense of agency in respect of sustained actions. Frankfurt's account, however, would make a mystery of how our sense of agency extends to instantaneous actions (intentional blinks, for example). Instantaneous actions do not occur for long enough to deviate from their proper path, and consequently guiding mechanisms of the sort Frankfurt proposes would have no role to play in respect to them. One either succeeds in blinking or one fails. No mid-action correction seems possible.[29]

By contrast, there are mechanisms that guide sustained actions but which can also operate in behaviour we might be doubtful about describing as acting. Say that Neil is knitting a scarf and knows that he must cast off after thirty-five stitches. Neil is an experienced knitter and is able to notch up stitches with an even and fluid hand. This would not have been so when he first learned to knit, and his developing the skill would partly have consisted in acquiring neural mechanisms that would guide his hand and

counteract the effects of tiredness, bad positioning, distraction and so on. We can also allow that those mechanisms would be necessary to his sense of agency about knitting skilfully. Suppose, however, that instead of casting off after thirty-five stitches, as he intended, he carried on to forty-five because he had begun to think about something else and had forgotten what he was doing. The mechanisms that ensure his steady knitting, we may say, would have continued to operate without his involvement ***qua*** agent. The stitches he carried out after his mind began to wander would not be thought of quite as actions, except in a very attenuated sense. Therefore, not only is it the case that guiding mechanisms are not necessary for all actions, but such mechanisms might arguably operate without actions taking place.

Frankfurt supports his claim that 'it is not essential to an action that it have an antecedent causal history of any particular kind'[30] with an example of a drug taker whose drug injection is caused by his addiction. Frankfurt concedes that the addict will have a desire to satisfy his craving for the drug and a belief about how that is to be done, which attitudinal conditions would render his action intentional. He argues, however, that these attitudinal conditions will affect the addict's behaviour by guiding it rather than by causing it. Frankfurt appears to think that what would cause the addict's behaviour would be an alien and non-attitudinal compulsion to take the drug. But this is surely mistaken. Suppose, for example, that someone has been repeatedly injected with heroin while he sleeps, and becomes an unknowing addict. If his 'supplier' were suddenly to discontinue the drug, he would experience a bodily craving for it without knowing it for what it is and with no idea about how to satisfy it. His symptoms would probably cause him to take to his bed or call the doctor rather than inject heroin. This points up a further problem with Frankfurt's account of the mechanics of drug-addicted behaviour. Frankfurt suggests that the addict's heroin injection will be caused by 'the compulsive force of his addiction', in the face of which 'he cannot help himself'.[31] A more plausible description of what causes the addict to inject the drug is that the addict has a conscious craving - which is impossible to ignore - and deeply uncomfortable symptoms, and that he takes the drug in order to satisfy the craving and be relieved of the symptoms. If this is what Frankfurt means when he says that the addict's behaviour is caused by the compulsive force of his addiction, then he has not offered us an example of causation by non-attitudinal conditions. If instead his claim is that the force of the addiction is somehow independent of the craving, the desire to be relieved of his symptoms and a belief about how that can be achieved, then the account is both implausible and otiose.

Nevertheless, Frankfurt's claim that the causal antecedents of actions are irrelevant to their status as actions is partly supported by how actions are experienced. When I raise my arm it does not seem as if the identity of the action is owed at all to any antecedent cause. There may be reasons ***for which*** I act, but ***that*** I act does not seem to derive directly from any event or state that precedes it. And yet if my sense that my actions involve or originate with me does not derive primarily from the contemporaneous operation of self-correcting mechanisms, what else could it derive from other than what causes them? While it is true, therefore, that whatever it is that enables us to discriminate actions from other bodily behaviour is not that they directly bear the traces of having been caused by earlier states or events of particular kinds, our sense that they originate with us might suggest that they are identified, perhaps indirectly, by means of their causes.

AGENT-CAUSATION

Attempting to do justice to the sense that our actions originate with us, and that our experiences of them are not such that they are caused by antecedent states or events, has been the motive for agent-causation theories of action.[32] Nagel's claim that the objective view of natural events is incompatible with the subjective perspective on actions as being one's own doing is shared by agent-causation theories. According to agent-causationists, however, since the way we view ourselves as agents is such an ineradicable feature of the way we think and behave, there is a presumption in favour of the truth of the agentive view. According to them, actions are caused by agents and the role of the agent in causing actions cannot be reduced to any natural event or sequence of events. Agent-causation may be a primitive feature of the way we picture and understand the world, on a par with event-causation in being both an understood and indispensable element of the way we see and behave, but unanalysable in terms of more simple or basic entities, properties or states. Agent-causation theorists, therefore, differ from those who claim that our ordinary view of agency is an illusion (or is likely to be) not in having a more positive or substantial theory of agent-causation, but simply in taking the ordinary view (as they see it) to be true.

According to agent-causationists, the causation of an event by an agent is not itself an event. Thus agent-causation is such that when an agent causes an event, although he causes it, there was nothing that he did to

cause it. Chisholm, following Aristotle, argues that each of us is an unmoved mover, such that 'In doing what we do, we cause certain events to happen and nothing - or no one - causes us to cause those events to happen.'[33]

All of this has led compatibilists and naturalistic philosophers to view libertarian theories of agent-causation as metaphysically fabulous and empty of substance. But as Chisholm has argued, since such theories claim no more than fidelity to a view of agency we all unreflectively share, it is incumbent on their critics to show either that this is not so or to give some idea about how we could come to do without that view of agency:[34] and this, despite some promising attempts, they have failed to do.

Setting aside charges of obscurity and fabulous metaphysics, does the concept of agent-causation do justice to one's sense of the difference between raising one's arm and one's arm just rising? It would be surprising if it did not, since it is in a way just the claim that there is no more to what goes on when one acts than one's sense of what goes on. And fidelity to our prereflective sense of acting is the principal recommendation of the concept of agent-causation. If we were to find reason to doubt that agent-causation does adequately capture our sense of acting, there would be little reason to consider it further.

It does seem that of the approaches so far considered, the concept of agent-causation best captures the way we experience acting. When I raise my arm, it does appear as if it is me that does it, rather than some antecedent state or event. It does not seem as if I have to do anything to bring it about: I just do it. My relation to the action of raising my arm is apparently one of a deep intimacy, which fails to register in causal accounts and which Frankfurt's theory about guiding mechanisms does not adequately explain.

Chisholm's view that the concept of action is such that we cause our actions to happen ***and*** that nothing causes us to do so[35] is not shared by Richard Taylor,[36] who argues that while it is true that the agent's role in relation to actions cannot be reduced or analysed in terms of events, it is not true that this precludes the agent's being caused to act as she does by some antecedent event.[37] While the first conjunct is enough for our conception of agency, the second is only required for our conception of free agency. Taylor describes a case in which a skier is caused by his fear of heights to perspire and to grip the seat of a ski-lift.[38] We would describe his gripping of the seat, unlike his perspiring, as an action, even though he is caused to do it by his fear of heights and could not refrain from doing it: our sense of the distinction presumably being owed to our understanding that there are certain circumstances that would cause the agent to

release his grip (say the possibility of receiving a punitive electric shock if he held on). We can say, however, that if we sometimes experience ourselves as caused to act by antecedent events, this is certainly not a feature of all actions, and causation of this kind will not contribute in any way to our sense of agency about basic actions.

We may wonder whether agent-causation is the most apt description of the agent's relation to her actions, since the term ***causation*** implies that actions are effects, whereas when I raise my arm it seems that I just do it. My relation to my actions is seemingly one of intimate involvement, rather than mere control or ownership or production of something distinct from me. Indeed my relationship to any action of mine seems more intimate than my relationships to anything else that is not essential to my being. My sense of agency, therefore, is not merely one of origination or possession (such as I might have to something I have made), but that I am essentially involved in my actions. The idea of agent-causation, even if it is a ***sui generis*** category, must have at least enough in common with the concept of causation proper to render the use of the common term meaningful. What it captures is the sense that somehow actions originate or issue from the agent. What it fails to capture is the sense that the agent is somehow ***inside*** her actions. The reader can test this out just by forming a fist: does it seem that you somehow gave a signal, or did something, and the fist formed in response, as it were? Or does it seem (as it does to me) that you simply formed and held your hand in a fist? It is not just that the idea of causation by the agent fails to capture this feature of our sense of agency, but rather that it excludes it. It does not seem, either, that changing the tag from ***agent-causation*** to some less controversial term such as ***agent-origination*** or ***agent-production*** will overcome the difficulty. The problem will still remain that all such terms imply a relationship between discrete items, whereas our sense of agency is such that we are involved in our actions for as long as they endure.[39] This is a defeating problem, of course, not just for agent-causation theories, but for standard causal analyses of action, which also attempt to account for actions in a way that treats them as temporarily discrete events and leaves our sense of intrinsic involvement in our actions a mystery.

The concept of agent-causation, then, like the other theories we have considered, does not do justice to the sense or experience we have of ourselves as acting. If we place this alongside its obscurity,[40] we have good reason to reject it as an account of what happens when someone acts, or even as a potential candidate for such an account. We come, then, to the conclusion that the concept of agent-causation, despite its initial promise, does not provide us with a defensible theory of action.

CONDITIONS OF A SATISFACTORY THEORY OF ACTION

The three types of theory we have examined so far have all failed to account for our sense of what goes on when we act. Their respective failures to do justice to, or at least to explain, our understanding of actions can be characterised as placing them on opposing horns of a dilemma. Our sense of acting is such that (1) our actions are thought to originate with us *qua* agents; and (2) we are intrinsically involved in them while they are taking place. Causal analyses of action and agent-causation theories account for (1) by reference to what it is that is thought to cause us to act. Such accounts, however, necessarily imply that the relation of actions to their causes is a relation between discrete items, which conflicts with (2). Frankfurt's account of actions as guided by causal mechanisms attempts to come to terms with (2) but fails to account for (1), since it denies that actions are in any way distinguished by what initiates them. Since it is hard to imagine how an explanation of our sense of the origin of actions can be given without reference to what it is that causes them, and that any account of actions as caused implies that they are related to their causes as discrete items, and since any such understanding of the relationship between actions and agents is contrary to our sense of actions as intimately and occurrently involving agents, (1) and (2) are *prima facie* incompatible.41 A satisfactory theory of actions, if there be such, must therefore overcome the apparent incompatibility between these two features of our sense of agency and be able to account for both.

It would be useful, before proceeding further, to summarise the conditions we have so far identified for a satisfactory theory of action. Beginning with the two horns of our dilemma, any satisfactory theory of actions must account for our sense that actions involve:

(I) origination by the agent;

(2) intimate and occurrent involvement of the agent;

to which we may add

(3) intentionality;

(4) agent control.

The theories we have considered so far have differing accounts of what is required for condition (3). According to Davidson, the content of an agent's intentions in acting is given by the attitudes that cause her to act and which

rationalise her doing so.[42] Agent-causation theorists argue that the intentions behind free actions are given by the agent's purposes in acting, but that the purpose of an agent's action, although it can explain it, does not cause it. Frankfurt has it that the intentionality of an action consists in its being purposive behaviour carried out by the agent, and the agent's contribution is given by the presence of causal mechanisms that guide her behaviour. I have suggested that Frankfurt's account, although insufficient as it stands, does meet the requirement of our sense of continued control and involvement in sustained actions, and such an account seems likewise indispensable to explaining the intentionality of sustained actions.

Condition (4) might be considered unnecessary, since it is supposedly our notion of control over our actions that (1) and (2) are meant to express. It is possible, however, that we might arrive at an account of actions that meets conditions (1) and (2) but in some way fails to satisfy the notion that we have control over our actions.

All the theories so far considered have failed to meet one or more of the above conditions, and, as I have suggested, there is no surprise about that since two of the conditions, (1) and (2), appear to be incompatible, or, at any rate, any attempt to satisfy condition (1) must, perforce, posit some originative source for actions, which in being distinguished from them will clash with condition (2). Any satisfactory theory must therefore overcome this seeming conflict between (1) and (2). If the conflict cannot be overcome, then of course we must concede that our sense of agency in respect of basic actions is an illusion. In that case, it will be illusory not because it is incompatible with determinism or natural causation, but because it is conceptually incoherent.

I will now attempt to show that our sense of agency, in respect of basic actions at least, is not an illusion. The account I will set out, which I believe can overcome the apparent incompatibility between conditions (1) and (2), is a causal analysis.

A DEFENSIBLE CAUSAL ANALYSIS OF ACTION

The central problems facing causal analyses of action are now familiar. They are, first, that any theory about the initiation of actions will treat them as having a discrete relationship to what it is that originates them and must therefore, apparently, conflict with condition (2). In addition, causal analyses claim that actions are initiated by attitudinal states such as beliefs and desires, whereas our sense of agency is such that we experience ***ourselves*** as originating actions ***in response*** to our attitudes. This

sense that our role as agents cannot be reduced to the causal effectiveness of our attitudes seems to be borne out by the inability of causal analyses to specify a law-like conditional that links attitudes to actions andwhich can exclude the possibility of attitudes failing to produce an action in the face of a belief that an action is worth performing, and even in the absence of overriding wants (as in the second of the three examples of agency I set out at the beginning of the chapter). Finally, there is the problem that behaviour can be caused by attitudes that would rationalise it, but via deviant causal paths.

The problem of framing an unobjectionable conditional statement linking attitudes to actions, which would be both universally valid and would specify the link in such a way as to exclude the possibility of deviant causal paths, is peculiar to causal analyses and additional to the need to satisfy the conditions required of any satisfactory theory of actions. Before addressing those conditions, therefore, I want to look first at the problem of causal sufficiency. For attitudes to be understood as causing actions we must have some plausible and coherent explanation of why it appears that they fail to do so on occasion, or we must have some other candidate for causing actions. In fact we can have both, if the other candidate functions as a causal intermediary between attitudes and actions. The alternative candidate we require, in that case, is a mechanism that can cause actions while itself being triggered by attitudes.

If we attend to the phenomenal character of acting and deciding, is it possible to discriminate such a mechanism? Consider this: Marci intentionally raised her arm. She did so to get the waiter's attention. Would Marci have experienced the operation of a causal intermediary - a trigger - which was caused by her motive for raising her arm, and which in tum caused her to raise her arm? Not if she experiences actions as the rest of us do. What she would have experienced is just that she raised her arm with the intention of getting the waiter's attention. The intention might be well accounted for by reference to her antecedent attitude, but her experience of raising her arm would not. **If,** as I have claimed, I raise my arm and the experience consists just in my doing so, it seems that it is my doing and that is all there is to it. I do not perceive any antecedent state or event as bringing about the raising of my arm, and nor do I experience myself as bringing it about in the sense that I contribute something antecedent to the action. I just raise my arm. And this is what is expressed by condition (2). What is there, then, that distinguishes the content of my experience of raising my arm from that of my arm rising involuntarily? Not, for sure, that it is preceded by an occurrent motive, since that could be succeeded by some mental or physical aberration that causes my arm

to rise without any input from me. And yet there is a strong intuition that I would know which is the case. Rising my arm has the 'actish quality',[43] in a way that my arm just rising involuntarily would not. The content of the actish quality when rising one's arm, however, seems to come to no more than experiencing oneself as raising one's arm. It appears, then, as if something enables us to make the distinction, or gives us a sense of it, but not something that we can gain any sight of through direct experience. What this suggests is an intermediary trigger that is not directly experienced, but which has a perceivable trace in our experiences of action.[44]

Let our hypothesis be, then, that actions are behaviour that is universally caused by triggering mechanisms, which themselves are caused by attitudes, and which cause us to experience the behaviour they cause as actions while not themselves being directly perceivable. We already have a collective name for such triggering mechanisms (one that has stubbornly maintained its footing in discussions about free will and agency, despite claims that it has no useful role to play): the Will.

As it stands, however, the Will, thus conceived, does not resolve our problem about causal sufficiency. The hypothesis about the Will claims a two-stage causal relation. If the operation of the Will is assumed to be causally sufficient for actions, we are no further forward with the relation between attitudes and Will. How can it be that a motive that would strongly recommend a particular action, and which an agent knows she can and should perform at a particular time, nevertheless fails, even in the absence of a countervailing motive, to produce it, or, as we would now say, fails to trigger her will?

Sometimes we fail to do what we have good reason to do because habit inclines us to do otherwise. If I want to get up and the room is warm and I have no noticeable desire to stay in bed, one explanation of why it still takes me half an hour to be up and about is that I have developed a habit of lying abed.[45] If someone smokes a cigarette, although she has no occurrent symptoms of craving, and she has good reason to expect (on past experience) that she won't enjoy a cigarette right now (in the way she does when she is satisfying a craving), but does so to 'fill out the time', or for some similarly half-hearted reason, then again, a plausible explanation of her behaviour is that she has formed a habit of smoking when, for example, she isn't involved in some task.[46]

My suggestion is that we sometimes fail to perform actions that we judge ourselves to have good reasons to perform, or fail to forbear from actions we have good reason not to perform, not because we have stronger countervailing reasons, but because we have habits of thought and action that incline us to do otherwise. Our all-things-considered best judgements

sometimes fail to trigger our wills, therefore, because our wills have been conditioned by training or habit to be more responsive to particular desires or to respond to perceived situations in certain ways. The hypothesis should therefore be expanded with the claim that the Will is a mechanism that is amenable to habituation, which is to say, it is such that it can be modified so as to be more responsive to particular states (including perceptions) or events and less responsive to others. Since it seems implausible that one mechanism would be responsive to a wide range of mental states or events, what we probably have is a system of triggering mechanisms, which we refer to collectively as the Will.[47] Thus what causes a person not to act in the way she wants, thinks best, wisest or right may be a badly habituated will. This does allow, however, as our ordinary view of bad habits would suggest, that agents sometimes have the freedom to overcome the effect of past conditioning by actively seeking to recondition their wills.[48]

If the hypothesis about the Will is incorporated in the conditional analysis I quoted earlier, of what happens when someone acts, we get the following:

> *CAA1. If A* wants ϕ and believes x-ing is a way to bring about ϕ[49] and that there is no better way to bring about ϕ, and *A* has no overriding want, and knows how to *x*, and is able to *x*, and has inferred from this that *x* is worth doing,[50] ***and has a will that has been conditioned or habituated so as to be appropriately***[51] ***triggered by her wanting and believing all of this, then A x's.***

Since actions are sometimes performed for their own sake, we should add that:

> *CAA2. If A* wants to *x*, and has no overriding want, and knows how to *x*, and is able to *x*, and has inferred from this that *x* is worth doing, and has a will that has been appropriately conditioned or habituated so as to be triggered by her wanting and believing all of this, ***then*** *A x's.*

The notion of *A* wanting ϕ or *x* would need to apply not only to situations in which *A* has an occurrent desire for ϕ, but also to cases of trained or habitual behaviour where a routine triggers the Will to cause an action that realises a standing desire that may not be an occurrent cause of what one does (like getting up and then going to the bathroom and brushing one's teeth). In cases such as these *A's* wanting ϕ or *x* would not be an immediate cause of her x-ing, although it would be realised by it. If we take S[d]: to stand for situations (times, places, particular kinds of event or

incident, etc.) with respect to which *A* has a standing desire to do *x*, we can say that:

> *CAA3. If A* perceives that *S*φ*s* and *A* knows how to *x*, and is able to *x*, and has no overriding want, and has a will that has been appropriately conditioned or habituated to be triggered by her perceiving that *S*φ*s* ***then A x's.***

Taken together *CAAI-3* cover all cases in which an agent may be said to do what she wants. They also provide us with a joint causal analysis of action that says enough about how a causally sufficient relation between attitudes and the Will could be established that would provide us with a legitimate stand-in for a law-like conditional. If an agent *A* wants φ and believes x-ing is a way to bring about φ and that there is no better way to bring about φ, and *A* has no overriding want, and knows how to *x*, and is able to *x*, but *A* does not *x*, her failure to do so will be as a result of her having a will that is not appropriately habituated or conditioned to be triggered by her wanting \$.

CAAJ-3 does, however, fall somewhere short of a full analysis of action, which is to say: an analysis of what actions ***are.*** For one thing, it says nothing about the relation between basic actions and those actions in which we do something by means of a basic action (like voting by raising one's arm).[52] It might also be thought that an analysis of the causes of action is something other than an analysis of what actions are. According to Ginet, it is preferable 'to have an account [of action] in which the mark of an action is intrinsic to it'.[53] Such an account, he thinks, will include the actish phenomenal quality as part of the intrinsic content of an action. I have suggested that the actish quality is the means by which we are able to distinguish actions. Whether Ginet is right in thinking that the actish quality is indispensable to what an action is will depend on whether it is possible for someone to act without her action being accompanied by the actish quality: whether, that is, it would be possible for someone to be ***action blind.*** To answer this would take me beyond the problems with which this chapter is concerned: namely, what it is about actions that distinguishes them from involuntary behaviour, how we as agents are able to make that distinction and whether our experience of the distinction is veridical. (I give reasons, below, for rejecting Ginet's claim that the actish quality alone, unaccompanied by an expected bodily movement, could signal or form part of the content of a mental act of ***trying*** to bring about such a movement.) I suggest that actions are distinguished from involuntary behaviour by their being caused by the Will, that we are able to make

the distinction by the operation of the Will being indirectly signalled to us via the actish quality, and that the normal correlation of causation by the Will with the actish quality renders our experiences of actions veridical. Therefore, although the hypothesis about the Will, as expressed in ***CAAJ-3,*** does not provide a full analysis of what actions are, it would be indispensable to a full analysis and it is sufficient for our present purposes. For the sake of brevity, I will continue to refer to ***CAAJ-3*** as a causal analysis of action (rather than 'an analysis of what distinguishes actions from involuntary behaviour').

The hypothesis that the Will is a mechanism whose operation is not directly experienced but indirectly signalled also enables us to resolve the problem of deviant internal causal chains. Any behaviour that is not caused by triggering the Will will not have been caused in the right way. And since we are able to discriminate behaviour that is caused by the Will, by means of its causing us to experience such behaviour as intimately involving us, we usually know immediately what behaviour has not been caused in this way.[54] The problem of internal deviancy arises from the fact that the putative attitudinal causes of action can cause an awareness of those attitudes, and that this awareness might cause behaviour that realises those attitudes by a deviant route. If we are not directly aware of the operation of the Will, and are only indirectly aware of it once an action has commenced, there is no possibility that awareness of the operation of the Will might cause behaviour by a deviant route.

WHY A CAUSAL ANALYSIS OF ACTION CANNOT INCLUDE INTENTIONS

Since I claim that the problem of internal causal deviancy is resolved by the hypothesis about the Will, it will be instructive to contrast it with an alternative approach to the problem. Several advocates of the causal theory of action have met the challenge of internal deviancy by attempting to specify a link between intention and behaviour in such a way as to exclude the possibility of deviant internal causal routes. Some have sought to do so by means of special categories of intention ('proximal', 'concurrent', 'final-stage'), which are said to be the immediate causes of actions, thereby excluding any causal space for deviancy;[55] others by specifying counterfactual or functional responsiveness of an action to the content of the intention that causes it and any salient features of the situation in which it occurs.[56]

The fundamental objection to such approaches is that they all violate the

requirement that a causal analysis of action must not include an item whose definition involves or presupposes the category it is supposed to analyse. The concept of intention necessarily presupposes the concept of action. No intention can be described without reference to an action or to an intended outcome of action,[57] and therefore intentions cannot form part of any general analysis of action.[58] It is a conceptual impossibility to intend any behaviour or outcome that does not involve or require an action (or intentional inaction).

Those who argue that actions are always caused by intentions can attempt to sidestep this objection by identifying intentions in terms that do not involve or presuppose the concept of action. Bishop, for example, argues that causal analyses can avoid the need for a definition of actions by showing how it is possible for actions to belong to the natural causal order.[59] What we intuitively recognise or experience as actions will be distinguished from involuntary behaviour by there being purely event-causal necessary and sufficient conditions for actions. Those conditions are provided by final-stage or basic intentions. Intentions are thought to be the appropriate kinds of causes for actions, since they rationalise actions in virtue of being action-directed. There is no need for a definition of intentions either (which would bring us back to the problem that intentions can only be defined in terms of actions, therefore rendering the analysis circular), Bishop thinks, provided that

> an agent's having an intention can ***ontologically be realized*** in a state of affairs that involves no agent-causal relations. Under a functionalist account of mental states, intentions can be distinguished from other sorts of mental states (wants, desires, etc.) according to the kind of functional role they play. And we can accommodate the idea that an intention involves the agent's setting himself or herself to achieve a goal simply by identifying intentions as those mental states that play a specially executive role - which are, so to say, 'functionally closer' to those event-types that initiate peripheral movement than any other kind of intentional state.ro

Bishop therefore suggests that intentions can be identified by their functional role in relation to the initiation of peripheral movement. Since the problem that a causal analysis of action must address is how to distinguish between those peripheral movements we count as actions and those we call involuntary, Bishop can only be referring to the former. A causal analysis of actions must show not only that actions have necessary and sufficient causes, however, but also that those causes are of the appropriately

rationalising kind, i.e. not only that they cause actions, but that they are the right kind of cause to ontologically realise 'the agent's setting himself or herself to achieve a goal'. Intentions are thought to meet this requirement in virtue of their action-directed propositional contents, but a functional account that identifies intentions solely by their executive role in causing actions is insufficient to identify them as having action-directed propositional contents and therefore insufficient to show that they can causally or functionally realise the agent's setting himself or herself to achieve a goal. It might be argued that, having identified intentions as discrete causal states by virtue of their functional role in relation to actions, we can rely on agent descriptions of these states to identify them as having action-directed propositional contents. But in the first place it is very doubtful that agents always experience their actions as preceded by such states (of which, more below), and in the second, this would involve identifying intentions in terms of the concept of action, and therefore we would be back to analysing actions in terms of causation by states whose definition or identification presupposes the concept of action. The most that Bishop's proposed functional analysis of the causation of actions could show is that there is a causal basis for our intuitive distinction between actions and involuntary behaviour. Without a further functional account of how the relevant action-causing states can count as intentional, our intuitive distinctions between actions (as behaviour that is under our control) and involuntary behaviour would be epiphenomena!. Bishop's attempt to develop a causal analysis of actions, without defining actions or the states that cause them, therefore fails and is inevitably forced back on identifying or defining intentions in a way that presupposes the concept of action.

It might be thought that the problem can be gotten around if we take the view that intentions, although conceptually dependent on actions, are reducible to conjunctions of beliefs and desires, since neither of these concepts presupposes that of action. For this to be accepted, however, at a minimum it would need to be shown that conjunctions of the appropriate types of beliefs and desires are materially equivalent to pre-action intentions. That in tum would require that good reasons be provided for thinking that no one could have the types of beliefs and desires in question, properly specified, without intending the corresponding action type. Since it is always logically possible that any belief and any desire can be overridden (without being negated) by other beliefs and desires, such an enterprise seems doomed from the start.

Analyses of actions that specify intentions (of whatever sort) as the basic causes of actions, belong to a family of mistaken approaches to the problem of action, including those volitional theories in which volitions

are defined as proximal intentions whose contents can generally be expressed thus: 'I shall do *A* here and now',[61] and others in which decisions play the key causal role; all of which employ concepts that assume the category they are supposed to explain.

It might be argued that any action-causing intentional states, beliefs and desires included, must be action-directed. My argument, however, is not that intentions are prevented from occupying the causal role in causal analyses of action because they are action-directed, but that they cannot do so because they cannot be ***defined*** other than in terms of actions and it would be circular, therefore, to define actions in terms of them. It is far from obvious, moreover, that all actions are caused by action-directed intentional states. Two examples should suffice to show this:

(a) 'Before either of us knew what we were doing we found ourselves kissing.'

(b) 'I had fully intended just to remonstrate with him, and to calmly explain to him how much harm his actions had caused, but as soon as I saw him, and before I even realised what I was doing, I was clutching his lapels and shouting in his face.'

In both examples the agents describe themselves as being taken by surprise by what they did and therefore as acting without prior intention. Against this it might be maintained that the accounts given could mask the existence of what would have been unconscious intentions to do what they did. I have no conclusive refutation to offer against this, but the contradiction of the agents' accounts seems both implausible and gratuitous and I can think of no reason to accept it other than to save those causal analyses of action which have, through lack of imagination, given states with action-directed propositional contents the generic causal role. (It may be that philosophical commitment to the idea that actions are always preceded or caused by intentions arises because when one is thinking about actions, and one's attention is focused on one's current actions, it is impossible for the actions on which one's attention is focused to be carried out without prior intention: such actions are, after all, carried out for the prior purpose of examining them. In that case, philosophers who believe that actions are always caused by intentions may have arrived at that conclusion by mistakenly generalising from a feature of actions that are performed in a peculiar context.) Moreover, if we were to accept such an account of what goes on in cases like the above, some account would then have to be given of why an unconscious intention should triumph over a

conscious one. The most obvious explanation would be that the agent's unconscious intention acquires its force from the feelings that cause it. But in that case some reason should be given for withholding Ockham's razor from the unconscious intention and just having the agent's strong feelings and the occasion doing all the causal work.

It might also be claimed that there could in both cases have been an immediate intention formed just before or concurrent with the action described. Again, there seems no obvious reason to accept this, and in addition the presence of such an intention would contradict the agents' own descriptions in the examples. The narrators in each example describe themselves as having been taken by surprise by what they did, which is incompatible with their having formed any prior intention to act; even an immediate one. Experiences such as this, moreover, ***are*** familiar to us. Since analyses of action as essentially involving causation by action-directed states are incompatible with such recognisable experiences, it is the analyses that should be abandoned.[62]

A solution of the problem of action requires an analysis that allows a causal role to all intentional states (including awareness of circumstances, as well as beliefs and desires), rather than just intentions. The resolution I have suggested is that actions are behaviour caused by mechanisms that are triggered immediately and exclusively by intentional states, that the differential responsiveness of these mechanisms is modified by habit, and that we are only indirectly aware of their operation through the actish phenomenal quality. And since, as I have already pointed out, the problem of internal deviancy arises because unmediated causation of actions by reasons allows the possibility that awareness of such reasons can bring about their causing actions by deviant routes, mediated causation via mechanisms whose operations we are not directly aware of contains all we need in order to exclude internal deviancy.[63]

Let us briefly consider one further argument in favour of the view that basic bodily actions are always preceded by volitions, which may appear to pose a problem for the hypothesis about the Will. According to the hypothesis, actions are always caused by the Will, from which it follows that the operation of the Will is not itself an action. Ginet has argued that if there is a brain process (such as the operation of the Will) specific to voluntary exertion, this process must be 'sufficient for the occurrence of a mental action [a volition], in which the subject tries (or wills) to cause her body to exert'.[64] This must be so, he thinks, because if the brain process occurs, but somehow fails to cause an exertion, 'then it must seem to the subject that she has at least tried to make the exertion.... But to try to act is to ***act.***'[65] But will the subject be right in thinking that she has

tried? I take it that trying to do something necessarily requires that we know how to try to do it, or that we have some idea of how we might try to do it. Thus if my legs are partially paralysed and I try to move them in the usual way, although I might not be able to describe how I usually succeed in getting them to move, it can be allowed that I do try since I succeed in getting them to move a little. If I cannot wiggle my ears and therefore don't really know how to do so, I might try to do it by contracting the muscles inside my ear, in the mistaken belief that it might have that effect. Or again, if my legs are completely paralysed, I might mentally focus on how I am normally able to move them in the hope that this will have the desired effect; and if this fails, I might rack my brains thinking of other things I could do to get them to move. Now suppose that having vainly gone through all of this I simply lie immobilised but very much wanting my legs to move, and that while I do so my will continues to issue unheeded signals that they should move: am I still trying? I might say that I am trying, and it would be callous of anyone to deny it, since I ***have*** tried and there is no doubt that I very much want my legs to move. But since I have no idea about how I might bring my legs to move, it would be true to say not only that I cannot move them, but that I cannot even try to do so because I do not know how to try. Someone in this situation might, for all we know, continue to be indirectly aware of the operation of her will if the actish quality were signalled to her without the usual accompanying motor and spatial sensations, but that would not be enough for her properly to count herself as trying to move her legs. The actish quality we experience when moving our legs normally signals that those movements are actions (and therefore caused by the operation of the Will): it does not signal some prior mental act. If one were to experience the actish quality without the usual kinaesthetic sensations, it would not signal an action, since there would be no action, but only that the process that usually causes an action has taken place without any action having resulted. I see no reason, therefore, to accept that basic bodily actions are necessarily caused by volitions, intentions or any kind of mental action.

RESOLUTION OF CONDITIONS (1) AND (2)

I suggested earlier that in addition to causing our actions the operation of the Will, while not itself being directly perceivable, causes us to perceive actions as our own doing. This is all we need to satisfy condition (2). If it does not seem that any antecedent state or event has brought about the raising of my arm, that is because the antecedent event that directly caused

it was not directly perceived. If it seems that it was me that raised my arm, although it does not seem as if any antecedent state or event caused it, that is because causation by the Will 'shows up' - is indirectly perceived - as the action's being my doing: as intimately involving me. What we experience, as I said earlier, is that 'we just do what we do'. Nothing other than a causal mechanism that is not directly perceived but whose operation is indirectly signalled could possibly account for this.

However, if there is an incompatibility between conditions (1) and (2), then as I said, a theory of action that can straightforwardly satisfy one of the conditions will face difficulties with the other. Condition (1) refers to our sense that our actions originate with us. I have argued already that causation by rationalising attitudes cannot occupy the agentive role we experience ourselves as having in originating actions, since we understand ourselves as acting in ***response*** to such attitudes. I have also suggested that if a person's will is not suitably conditioned, she may have appropriately rationalising attitudes and no conflicting wants etc. and yet fail to act. If our sense of actions as originating with us does not derive from rationalising attitudes, however, it does not seem as if the Will can be responsible for this sense either. If we do not directly experience the causation of actions by the Will, and are aware of it only indirectly in the sense of intimate involvement in actions, it is unlikely that it can account for our sense of the origination of actions. Nor is there any other feature of the Will that might obviously account for this, since its identity, inasmuch as it is relevant to philosophical questions about agency, comes to no more than its causal role in respect of attitudes and actions.

Let us look more closely at what is called for by condition (1). We have a sense that our actions somehow originate with us. That sense does not derive from our experiencing either attitudes or the Will as causing us to act. Indeed it cannot be thought to derive from the bare experience of acting at all, since as I have mentioned, when one performs basic actions like raising one's arm, it appears as if one just does so and that one's doing so is not owed to the occurrence of some antecedent event by means of which one does so. To say this is no more than to re-state condition (2), and therefore, seemingly, to exclude any account that could satisfy condition (1). However, if the sense of origination does not result from any direct experience of what causes actions, it may have another source.

In Chapter 4 I touched on the soft-determinist argument that determinism is not only compatible with moral responsibility but positively requires it. According to that argument, actions that do not issue from stable virtues or vices of character would be random and capricious and therefore not the kinds of happenings for which anyone could properly be held

responsible. Without assenting to the claimed dependence of moral responsibility and agency on determinism, it can rightly be argued that the account points up the importance to our sense of agency of some idea of ***character*** as the source of actions or as guiding actions. We may consider ourselves as able (in some sense) to carry out either of two opposing actions, and we may sometimes be unsure of how we will act, but there is, nevertheless, a sense of actions as originating (even if only partly so) in something enduring and stable. This sense is expressed in the idea that actions issue from, or reflect, something we describe as ***character.*** What we take character to be is often read off from how we have acted or how we have tended to act in past situations.

My suggestion is that what we have in mind with character as the source of action partly includes the Will.[66] The Will, according to the account I have given, can be habituated to be more responsive to some attitudes than others, for good or ill, and can be habituated by favourable motives, intelligence, circumstances and so forth. What else could character as a source of action refer to, therefore, other than the characteristic ways in which the Will is triggered to act in response to certain attitudes and situations? I am arguing that the notion of a relation of origination between character and actions is common to all of us and is the basic source of our sense of actions as originating with us. It is therefore important to stress that this sense of the role of character is not one of character rigidly fixing us to certain types of action. The view of character I am advocating is simply that in the influence of habit on the Will there is a stable and enduring influence on actions, which need not determine our actions in every detail. Hence, the sense of actions as originating with agents derives from the notion that actions have a (partial) source in one's character. ***Character*** partly refers to the way the Will is attuned to various attitudes, and this is affected not just by past actions, but by what practical reasoning reveals about our attitudes and the best ways to realise them. From what we know of our past actions, therefore, we are able to develop a sense of actions as originating in something stable within us or about us, and also that that something responds to attitudes and practical reasoning. That something is the effects on the Will of past behaviour and training.

There is, I think, a further source of the sense that our actions originate with us, which is that our actions are often clearly directed towards what we want and in line with our reasonings and judgements. If our experiences of actions, ***qua*** actions, do not involve their being caused by antecedent attitudes, we do experience them, nevertheless, as often responsive to such attitudes. Since our attitudes are ***our*** attitudes, the responsiveness

of our actions to our attitudes is a further source of the sense we have that our actions somehow originate with us.

The seeming incompatibility between conditions (1) and (2) can now be resolved. In the first place, the sense we have of actions as originating with us is different in kind from the sense we have of being intimately involved in our actions. The sense we have of actions as originating with us is not, after all, one of a relationship between our actions and ourselves ***qua*** agents, but has its source in the relationship between our actions and something within us, and also in the responsiveness of our actions to what we want. We do not experience ourselves as originators of actions in the same way that we experience ourselves as intimately involved in actions. The sense we have of ourselves as originating actions is something we infer from our attitudes and from past actions, rather than something we directly experience. This corresponds to the experience of there being no antecedent state or event that is required for lifting one's arm etc.

INTENTIONALITY

According to Davidson, as we have seen, a person acts with an intention if the attitudes that cause her to act are such as to rationalise her acting in that way. An alternative approach would be to treat intentions as ***sui generis*** states that are independent of one's behaviour or what causes it. While a case can be made for this in respect of prior ***intending to,***[67] there seems no need for an additional category to explain the intentionality of actions.

Consider, for example, what happens when the intention with which someone acts changes in mid-action. Suppose that Tom, while out in the park, sees an elderly man being robbed and assaulted by two men, mistakenly identifies the victim as the very wealthy Mr Megabucks and rushes to his assistance (his hope of a fat reward having overcome his fear of being set upon by the two muggers). As he gets closer he realises that this is not Megabucks, and, consequently, that he has nothing to gain (and much to lose) by helping him. At the same time, however, he is seized with a feeling of self-revulsion at his greed and lack of concern for the old man, and therefore continues to come to his aid.

The switch in Tom's intentions from acting for gain to acting just to help someone in distress is straightforwardly accounted for in terms of what causes him to act. To begin with, his greed and his belief about a possible reward cause him (trigger his will to cause him) to rush to the old man's assistance. The realisation that he has misidentified the old man creates a change in his attitude, which, taken together with a desire for

self-preservation and the danger of intervening, should cause him to stop what he is doing. But his new attitude of self revulsion and the consequent desire to help someone in distress cause him to continue. It is difficult to see what role intentions in acting, as a *sui generis* category, could have in the explanation of Tom's actions or the change of intentions with which he acts. Any causal role in relation to actions that could be claimed for intentions in acting (rather than intentions to act) would be gratuitous since we already have an adequate causal account.

Explaining intentionality with reference to what causes us to act is consistent with the experience of sometimes being unsure why one is doing something, either because one has forgotten why one began doing it or because one has overdetermining motives. To know why we act as we do is to be aware of our prior motives for acting, and to infer that these are the reasons for which we act. It is a common enough experience (if not an everyday one) to realise in a moment of insight that one's reasons for acting are other than one took them to be. Any account of intentions as *sui generis* states would have to explain this phenomenon by claiming that we do not always have direct access to our intentions. But this creates a mystery where none previously existed. We should then need to know what it is about intentions that sometimes prevents us from knowing them or being clear about them. Why, in particular, if intentions are *sui generis* mental states, should they be rendered less accessible to us whenever we have overdetermining prior motives? The best explanation of why we are unsure about our intentions in such situations is that we are unsure what motives have caused us to act.

A key problem for the idea that intentions are given by the attitudes that cause an action is that it owes its plausibility to the notion that the attitudes that cause actions also rationalise them. The relevant sense of *rationalise,* however, calls for something more than the bare fact that a particular attitude an agent has provides a reason for his acting in a certain way. Say that Ian is sitting in a diner waiting for his lunch to be served. He is hungry and tired of waiting, and he knows he can satisfy his hunger by snatching the plate of the person sitting next to him. This gives him a reason for snatching the plate, but he would never dream of doing any such thing. We may say, then, that, given other attitudes he has, snatching the plate would not be a rational thing for him to do, or that being hungry and knowing that he can satisfy his hunger by snatching the plate do not give him sufficient reason to do so. It would follow from this that an

action is rationalised by the attitudes that cause it (or cause the Will to cause it), if those attitudes, taken in conjunction with other relevant attitudes (about propriety, etiquette or whatever), provide an agent with sufficient reason for doing it. Several philosophers have thought that this requirement is expressed in the principle that an agent acts on a judgement about what is the best thing to do.[68] But agents do not always perform what they rationally judge to be the best thing to do, and we do not consider actions that are weak-willed to lack intentions. The problem arises from the assumption that the strongest reasons are, or ought to be, the strongest causes.[69] The assumption is required, seemingly, because without it there is nothing inherently rational in the initiation of actions and no explanation of why one reason rather than another causes us to act.

If I lie in bed while desiring to get up, we might describe what keeps me in bed as apathy (understood as a state of mind). But apathy implies a lack of desire to be up and about, whereas I have suggested that it is possible to lie abed while desiring on balance to get up. The temptation for causal analyses of agency will always be to deny that such cases of weak-willed behaviour actually do involve the agent doing other than what she wants to do. It seems to me, however, as it has seemed to others, not only that this is possible, but that it is something we often experience.

What has been lacking in causal analyses of action is any consideration that the causal strength of reasons for acting is not equivalent to their *prima facie* rational strength, but is affected by the medium on which they act. As the children's game 'scissors and stones'[70] indicates, the nature and structure of the medium that is acted upon makes a difference to the effectiveness of that which acts on it. I argue that the medium on which reasons operate is the Will, and that the responsiveness of the Will is not simply attuned to the rational strength of reasons, or intensities of desire, but is modified by past behaviour and may be triggered by perceptions as well as attitudes. When I lie in bed, therefore, although what I want is to get up, what prevents me from doing so is that my past behaviour has established a habit of lying in bed which has had the effect of making my will less responsive to my desire to be up and about.

The causally strongest reasons, therefore, are not the ones that it is most rational to act on, but those that are constituted by whatever intentional states one's will is most attuned to. The Will is partly attuned to what one wants and how much one wants it, but is modified by one's past behaviour. An agent's intention in acting, therefore, is given by the reason that triggers her will to act; and the reason that triggers her will to act may be represented, as in ***CAA3,*** by a perception of a situation as being of a type

that figures in a standing desire (taken together with beliefs that the agent knows how to *x*; is able to *x*; etc.).

What then becomes of the claim that intention in acting is given by the attitudes that cause an action also being the ones that rationalise it? If an agent acts for reasons she judges not to be the best, then they cannot be said to rationalise what she does in the sense of providing her with sufficient reason for doing what she does. If we say that an agent's intentions in acting are given by the reason that triggers her will to cause her to act, and leave it at that, we will not be able to say that causation of actions is rational in any respect since it would not reflect or embody any judgement but would be simply a matter of how the Will happens to respond to particular attitudes and situations. If that were so, it would also undermine any notion of control by the agent over her actions. I think that the way out of the difficulty is to say that the final judgement about what to do is given by the Will, and that this judgement is constrained by the way in which the agent's intentional states, including practical judgements, stand in relation to her existing habits and training.[71] Such a compromise between reasoning and the influence of past behaviour should not be thought of as only partly rational on account of the limitation on the influence of practical reasoning. Since practical reasoning can be carried out irrationally, it is possible that irrational consequences of practical reasoning can be curtailed by the influence of past behaviour. The idea that it is the Will that passes final judgement on reasons, and that they do not simply select themselves according to their respective strengths, is in line with the intuitions of libertarians from Reid onwards that reasons are effective only because the agent gives effect to them.[72]

The notion that intention in acting is given by the attitudes that cause us to act also being those that rationalise our acting can therefore be retained. We can retain it if what causes the agent to act is determined by practical reflection and by the influence of past behaviour, culminating in a 'judgement' about what to do (rather than what it is best to do).[73] An agent may regard such 'judgements' as mistaken, because she is alienated from her past behaviour and its continuing influence on what she does, and this possibility will be examined in the following chapter. But such judgements will be rational in the sense that they are arrived at through a principled process of action selection.

Having arrived at a satisfactory account of what it is to act with an intention, a problem remains about whether it is possible to act without an

intention. Suppose someone has reasons for x-ing and that these are also reasons for not y-ing, and she is aware of no reason why she should y, but she y's anyway, and does so knowingly, then it seems true to say that she y's unintentionally. A person might, for example, be caused to spit at anyone who offers her a compliment; because of some long-buried trauma, perhaps. Her will, in that case, would have been conditioned to cause her to do things she has no reason for doing. It might be argued that neurotic behaviour of this sort is not really acting, but I can think of no reason for accepting this. If it is correct to describe such behaviour as acting without intention, then *CAAJ-3* will need to be complemented by an analysis of what happens when someone acts without intention. If S_{ix} stands for any situation the perception of which is followed by an unintentional action on the part of the perceiver, then we may say that:

> *CAA WI. If A* perceives that S_{ix}, and *A* does not want to *x*, and does not believe that x-ing is a way of bringing about anything she does want, and knows how to *x*, and is able to *x*, and has a will that has been conditioned or habituated to be triggered by her perceiving that S_{ix} so as to cause *x*, ***then*** *A* x's unintentionally.

Taken together, *CAAJ-3* and *CAA WI* should cover all cases of basic action.

CONTROL

There is no straightforward answer to whether our account of the Will can satisfy condition (4), and this is simply a reflection of the ambiguity about what we mean by ***control*** in respect of simple actions. If we take ***control*** to mean 'moral responsibility entailing', then the account does not satisfy condition (4) on any plausible interpretation of what is required for moral responsibility. However, the sense of control we wish to satisfy is not that which is sufficient for moral responsibility, but rather that which is embedded in the distinction between voluntary and involuntary behaviour.

Although we can distinguish ourselves as agents from our attitudes (even to the point of having strong negative evaluations of such attitudes), as I said earlier, our attitudes are ***our*** attitudes. Our wills are responsive to our attitudes. The differential responsiveness of the Will to our various attitudes, routines and intentions is affected by the varying intensities of our desires, by practical reasoning and by prior conditioning by past behaviour. The degree to which responsiveness of the Will is affected by our

past actions seems to be incompatible with any sense of immediate control, since past actions are beyond our control. This might lead us to conclude that to the degree that a person's will is responsive to the varying intensities of her wants, unmediated by past conditioning, she does what she wants and is therefore in control. On the other hand, a person who values moral behaviour but has a strong desire to do something immoral and self-serving might rely on her habituated sense of right and wrong to overcome that desire, and feel herself to be in control just because her will has been conditioned to be more responsive to her moral sense than to her occurrently strongest desire. This is in line with the conception of freedom as doing what we really want because it is what we really want, and anticipates what I have to say in the following chapter about fully human agency. For now I make the point that while our understanding of control over actions may be anchored in the responsiveness of the Will to the intensity of one's desires, it is not inimical to the way in which that responsiveness can be modified by one's past behaviour.

Does this mean, as Double would argue, that our sense of agent-control is vague and multireferential, even incoherent? Say that what Karen most desires is to embark on a wild love affair with someone to whom she is deeply attracted, but she doesn't do that because her will is more responsive to a sense of duty and loyalty: should we think of her as more or less in control than if she were to act on the desire for an affair? If she feels frustrated and enslaved by her sense of duty, then she is likely to feel that if she could do what she most desires she would be more in control. But if she feels confident and secure in what she does (although disappointed), she will also feel that her ability to do the right thing places her in control. Her sense of control, therefore, will depend on whether she ***identifies*** with what she does. What it means to identify with one's actions is dealt with in the following chapter, but for now we may note that the dilemma about whether control is given by responsiveness of one's will to the intensity of one's wants, or to the standards embodied in one's practices and habits and established by one's past actions, arises from the relative motivational complexity of human beings and their capacity for self-conscious reflection and evaluation. Lambs and lions, in true Hobbesian fashion, pursue whatever they most desire and avoid whatever they most fear. They lack the capacity to reflect on their wants and aversions. What it means to describe them as being in control of their behaviour is fixed by their motivational system, which, we suppose, is differentially responsive to the intensities of their appetites and aversions. Human beings, having a motivational system of greater complexity, including reflective or higher-order desires, and values, can have access to a higher or more complex

level of control than that of animals. But since what is required for the sort of control we believe to be peculiarly human is different from what is required for animals, we have conflicting intuitions about what is required for control.

For now we can say that since the additional complexity of human motivation does not replace our simpler animal motivational system, we can be thought of as having control at least in the way animals do, provided, again, that our wills are differentially responsive (in some degree) to the intensity of our wants and aversions. Even where someone's will is not responsive, in some situations, to what she most wants, provided that it responds in some degree to the intensity of her wants and aversions, she can be said by that to have some basic control over her actions.

DUAL-CONTROL AND DUAL-RATIONALITY

Another attractive recommendation of the hypothesis about the Will is that, taken together with the failure of bivalence in respect of the agent's position in relation to statements about difficult decisions she has yet to make, it shows how we can have the kinds of dual-control and dual-rationality (to be able to choose either way, and to be rational and in control however one chooses), which have figured significantly in recent discussions of free will. The argument against bivalence shows that in respect of any future difficult decision of mine I can categorically decide either way. Compatibilists have objected that, with everything just as it was at the moment of decision, if I could have decided either way the decision would have been random, and if that is so it could not have been either rational (since lacking a principled relationship to my reasons or judgements) or truly under my control. I would argue that the true force of the incompatibilist case for dual-rationality and dual-control is in the idea that we should have sufficient distance from our reasons for acting to be warranted in saying that we choose or decide ***for*** reasons without reasons themselves being the choosers.[74] If our reasons do not, without mediation, cause us to decide, but instead the Will is triggered to respond to reasons according to their respective strengths in relation to our existing dispositions and habits, then we can say that the functioning of the Will is such that it independently ***selects*** from among reasons. This independence is relative, of course, because selection is partly dependent on strengths of desires and so forth. But this, I think, is all the independence the Will needs in order to satisfy our sense of independence from reasons and motives.

Compatibilists may argue that this account fails to satisfy the requirement of dual-rationality, just because the Will is independent enough not to be activated by what one judges to be one's all-things-considered best reasons for deciding or acting. But since it is envisaged that the Will is conditioned in a principled or regular way by dispositions and habits, and since our judgements about our all-things-considered best reasons for deciding and acting can sometimes be mistaken, it is possible for it sometimes to be triggered in a way that, although contrary to what one judges to be one's all-things-considered best reasons for acting, nevertheless realises what are really one's best reasons for acting. If one can categorically decide either way between what one wrongly judges to be one's all-things-considered best reasons for acting and what are in fact one's best reasons, one can rationally decide either way.

The argument might still be resisted on the grounds that if the senses in which the two options are thought to be rational are not the same, what I have described would not be genuine dual-rationality: that one of the options would not genuinely be rational. But from the point of view of the agent, what she judges to be her all-things-considered best reasons for deciding would *be* the rational ones to act on. At the same time, however, knowing that her judgement is fallible, she can sometimes also consider that the alternative might be a more rational choice (whether or not it is), and if she has some idea or feeling about the considerations that might make it a better choice, then she will have reasons enough in respect of either alternative to make it rational for her to decide either way, i.e. she can sometimes decide contrary to her best judgement, knowing that there may be good reasons for doing so.

DEMONS AND MANIPULATORS

Many readers will be familiar with those perverse neuroscientists and demons whose manipulations so thickly populate the pages of books and articles on free will and agency. What would happen if one such were to modify some of the mechanisms that are my will so as to make them responsive to his instructions rather than to any motive of mine? Since I would not be acting for any reason of mine, I could not understand myself as acting with any intention: but could I continue to think of myself as acting at all? If I could not, then we would have to accept that causation by the Will is not enough for the behaviour it causes to count as acting.

If an agent's will were modified so that it could be triggered by the instructions of another, it would no longer be differentially responsive to

that agent's intentional states. Such modification would therefore alter the identity of the will that is modified. In that case what it would cause would, in one essential respect, no longer be actions, since they would no longer be caused by mechanisms that are triggered exclusively by the agent's intentional states. But since an agent's awareness of her actions as hers is caused by the operation of the Will being indirectly signalled to her, it is conceivable that modified wills might continue to have the same actish phenomenal quality. Therefore, although behaviour caused by modified wills would not be actions in the strict sense, it is not incompatible with the account I have given that it would still ***feel*** like acting.

If we assume to begin with that an agent knows nothing of these demonic modifications to her will, and that she is not directly aware of the manipulator's instructions, then she might find herself doing things she did not want to do and which she had no prior intention or thought of doing.[75] That in itself would not be enough to destroy her sense that it is she who has done these things. Certainly one would not identify with one's 'actions' if they were caused in this way, but identification with one's actions is a problem for our distinctively and fully human sense of agency rather than our basic sense of agency. One can act in the basic sense we have been considering, even if one fails to identify with one's actions.

Suppose that the manipulator were to instruct a person's will to cause her arm to rise: would she be able to distinguish the experience of this from her usual experiences of raising her arm? The answer, surely, is that there can be nothing in the experience of the raising of her arm itself by means of which she could make such a distinction. What would be different would be the lack of any antecedent motive or intention to raise her arm. That might lead her to wonder why she had raised her arm, but not to doubt that she had raised it. Experiences of this sort are not unfamiliar. As I have said, people do sometimes find themselves wondering why they did what they did - not because they had overdetermining motives, but because they are unaware of having had any motives for what they did - and deciding, perhaps, that what they did was an aberration, or that there must have been some unconscious motive, without being at all inclined to conclude that what they did was not really their own doing (maybe even wishing that they could believe that it wasn't).

If the manipulator were to keep his activities up for any length of time, so that his victim could never know what she would do next, it is unlikely that she could continue to think of her behaviour as her doing. This would not happen, as such, because the erratic nature of her actions would lead her to conclude that they were not truly under her control, but rather that since a person's sense of agency and control is built around her actions

having some degree of regular and intelligible connections with what she wants, her sense of self and agency would be undermined. It would not be that her actions would come to lack the quality of seeming to be her doing, but that the notion of ***anything*** being her doing, and even of there being a person to do it, would be undermined.

We would expect the same effect, only sooner, if the manipulator's instructions were all known to his victim. She would find the sense of herself as being able, even in principle, to affect the course of events slipping away from her. To begin with, however, the experience of raising her arm whenever the manipulator told her to, even though she did not want to, and even though she may have formed an intention to refuse him, would be akin to the experience of being compelled to do what one does, by fear, say, or like domination of a weak personality by a strong one. In other words, she would experience herself as involved in what she did but also as being caused to do it by something alien.[76]

If someone's will were made responsive to the instructions of a manipulator and entirely divorced from her motives, we would not regard any resulting actions of hers, whether or not she was aware of the manipulation, as free. That our account of the Will is compatible with manipulation of it, and therefore that our account of the causation of actions is compatible with their being unfree, has already been conceded.

CONCLUSION

We have now come to a causal hypothesis about basic actions that satisfies the four conditions set out earlier, and which overcomes the problems about causal sufficiency and deviant causal chains faced by existing causal analyses. In satisfying condition (2), moreover, the hypothesis is able to meet the objections of agent-causation theorists that our sense of agency and our experiences of actions are such that they are carried out by ourselves as agents rather than caused by antecedent events or states. We do not experience our actions as caused by antecedent events because we do not directly experience the operation of the mechanisms that cause them. Causation by the Will is signalled to us, however, in the phenomenal sense we have of actions as our doing. This also meets Frankfurt's objection that causal analyses of action direct attention away from the intrinsic content of actions. The hypothesis about the Will claims that what causes actions is also responsible for their being experienced in a particular way. Finally, the hypothesis, taken together with the logical indeterminacy (from the agent's point of view) of future difficult decisions, enables a positive

account of dual-rationality and dual-control, which, since it does not involve causal indeterminacy, evades the standard objections to incompatibilist accounts.

Having arrived at an account of actions and acting that does justice to the way we experience and understand the distinction between actions and involuntary behaviour, we can now proceed to an account of how non-illusory free agency is possible.

6 Free Agency

If the hypothesis about the causation of actions by the Will is correct, it will be as true of actions caused by drug dependency and kleptomania as those of a free agent. Unfree actions are ones from which an agent can feel *alienated,* such that in some further sense they are not really hers at all. Moreover, as an account of what distinguishes basic actions from involuntary behaviour the hypothesis is insufficient to account for a more complex relationship we take ourselves to be capable of having to our motives, deliberations and actions. It is not that actions must attain some putative level of complexity in order to be free, since we have no reason to doubt that many simple and straightforward actions are free, but rather that our being free depends on our sometimes being able to have this more complex relationship to our actions. What then does this relationship involve?

An engagement falls through leaving Carol with a choice about what to do next Saturday. What she would like to do is to get up late, do some house repairs and spend the evening watching television. But she is nagged by a feeling that what she ought to do is travel 150 miles to see Alex, who can't get around very well anymore and doesn't have many visitors. In the past she has been able to justify her failure to make the journey on account of her busy timetable, but now she mentally replays her mother's voice telling her what a pleasure it would be for Alex - and how little trouble for her - if she were to see him for just one afternoon. Suppose we say that the two motives - to do what she wants or to do her family duty - do battle for a time, and eventually the stronger motive triggers her will, thus forming an intention to visit Alex; and that when Saturday comes, the intention likewise triggers her will so that she gets into her car and makes the journey to see Alex. Such would be a Hobbesian account of her decision and subsequent action: only now with what it had formerly lacked, which is to say, it instantiates a causally sufficient relation between the Will and actions, is incompatible with deviant causation of behaviour and has in it that which accounts for the phenomenological character of actions as the doings of agents. But this, assuredly, is not how we take ourselves always to stand in relation to our motives, intentions and actions.

When I have conflicting motives, my relation to them is not (or not always) that of a passive bystander waiting to find out which of the combatants will prevail and therefore come to be the cause of my behaviour. Indeed, the notion that a conflict between motives somehow involves a process of literal struggle between them, although picturesque, is misleading.[1]

My motives can have contents that cbnflict in the sense of prescribing contrary courses of action, but otherwise any struggle that takes place is ***my*** struggle to decide in favour of one motive and against others; or as Reid puts it: 'Contrary motives may very properly be compared to advocates pleading the opposite sides of a cause at the bar'.[2]

According to our 'full-blooded conception of agency',[3] when confronted by conflicting reasons for acting we actively deliberate about what to do, and decide in favour of one motive, and against another, by forming an intention to carry out what it prescribes.[4] Thereafter we carry out the intention by acting. We do not, it is true, always understand ourselves to act according to such robust standards. We sometimes act half-heartedly and we sometimes do things under compulsion. If I hand over my wallet to a man who holds a gun to my temple, my actions are caused by fear and there may be no way I could bring myself to do otherwise than I do. But actions like this are not human actions ***par excellence.***[5]

If my actions are coerced by threats, or are in some way half-hearted, I may think of myself as acting unfreely. But if an action of mine is unfree, this means not only that it fails to meet a standard for free actions, but also that in some sense the action is not really mine: it is not what I really want to do. Of course, any action I perform is mine in the basic sense just by being an action that I perform, which is to say, it is mine because it is caused by my will. And if an action is caused by my will, then it will be experienced by me as my doing. But if I hand over my wallet to a man with a gun, I will not act in the fullest sense, because I am fulfilling his wishes. I may have chosen to hand over the wallet, but this is not choosing in anything but an ephemeral and insubstantial sense, since, as I argued in Chapter 3, the alternative to choosing the way I do is not a choice I could reasonably be expected to make. What then is the difference between a choice such as this and one of which we could say that it is me in the fullest sense who does the choosing?

An obvious answer is that since the alternative is not one I could reasonably be expected to choose, the choice I would make if someone were holding a gun to my head would not really be mine. Should we say, then, that a person acts or chooses in the fullest sense only when there are no consequences attaching to the alternative(s) that would render it too unpalatable for her reasonably to be expected to make it? The answer must be no, since there are cases in which people choose to reject options with bad consequences without our being inclined to say that the choices were not fully theirs. Someone suffering from terminal cancer, for example, might choose to die with dignity rather than suffer a prolonged end filled with pain or drugged stupefaction, and there is reason to believe that such

decisions are often sane, based on careful reflection and fully the agent's own. To this it might be objected that in such cases the sufferer would rather not be faced with such a choice; that she would prefer to have good health and long life. But as I pointed out in Chapter 3, all choices are based on circumstance, and it is always possible to wish for better circumstances than those we have. **If** we are to draw the line between full and lesser agency according to whether a person would ideally have chosen to face the choice she does, then few choices or actions will ever count as fully the agent's own.

If there is nothing to be found in the options themselves, or their consequences, to ground our distinction between half-hearted or coerced actions and those which are fully our own, we must look instead to ourselves and our attitudes about our actions and choices. What we want, it seems, is true self-determination (to be able to do what we really want because it is what we really want) such that we can say of the sequence of mental events that produces an action, that they are sufficient for the action to be described as fully the agent's own. What we want is a theory of free agency that completes the account of self-determination given in Chapter 3.

SELF-DETERMINATION AND IDENTIFICATION

Our conception of full-blooded agency is such that we view ourselves as able to evaluate and make judgements about, or selections from, conflicting motives,[6] and that we are able to put such judgements into effect by acting. The idea is one in which the agent occupies a distinct position in relation to her various desires and beliefs and is able to reflect critically on them and act independently of them. This is what I described at the beginning of the previous chapter in the third paradigm case of acting. The question I raised there was whether the picture of an independent agent, able to reflect critically and independently on her conflicting motives and to act on whatever decision she makes about them, is illusory; and whether, instead, actions are simply caused by the strongest motive (or, as we would now say, the motive to which the Will is most strongly attuned). Should the latter be true, decisions might be made between genuinely open alternatives in the way set out in Chapter 4, and actions might be directed at what one really wants because it is what one really wants, but what we do will still fall short of agency in the fullest sense. What we come back to is the idea that if actions are just the end products of causal sequences, even causal sequences as complex and differentiated as those described in the previous chapter, this may be all right for the beasts, but it fails

somehow to find a place for the kinds of intervention between motives and action we take human agents to be capable of.

What is wrong with standard Hobbesian accounts of the causation of actions, according to Velleman, is that

> Various roles that are actually played by the agent himself in the history of a full-blooded action are not played by anything in the [standard causal] story, or are played by psychological elements whose participation is not equivalent to his. In a full-blooded action, an intention is formed by the agent himself, not by his reasons for acting. Reasons affect his intention by influencing him to form it, but they thus affect his intention by affecting him first. And the agent then moves his limbs in execution of his intention; his intention doesn't move his limbs by itself. The agent thus has at least two roles to play: he forms an intention under the influence of reasons for acting, and he produces behaviour pursuant to that intention.[7]

As it stands, however, Velleman's account of what is lacking in the standard causal story - something that occupies the role of agent by being influenced by reasons to form an intention, and thereafter by executing that intention - would be satisfied by the Will. Velleman cites examples of causal deviancy and drug-addicted behaviour to illustrate what is missing from the causal story. Quoting Frankfurt, he suggests that an addict who takes the drug he craves may be 'a helpless bystander to the forces that move him'.[8] But if a drug addict is a ***helpless bystander*** in relation to what he does, it is in a different sense from that in which this may be said of someone whose behaviour is a product of causal deviancy, since the actions of the addict, no matter how reluctant, are ***actions,*** which is not true of behaviour that is produced by deviant causal routes. Velleman misidentifies the problem of what is required for full-blooded agency with the problem of basic action. What is required for full-blooded agency is not, as such, a state, event or entity that can occupy the agent's intermediate causal role between motives and intentions, and between intentions and actions, but (as Velleman goes on to say) one that can realise the position of critical detachment the agent is thought to occupy in relation to his motives and his ability to add decisively to the force of some motives rather than others by throwing his weight behind them.[9]

Finding a place in the initiation of actions for fully self-determining agents has been the goal of the hierarchical compatibilist accounts of agency I mentioned in Chapter 3, beginning with Frankfurt's 'Freedom of

the Will and the Concept of a Person'.[10] Frankfurt suggests that someone's behaviour is compulsive, coerced or otherwise half-hearted, if she is alienated from the motives that cause it. He argues that an agent participates in her behaviour by ***identifying*** with the motive that actuates it.[11] The problem of distinguishing half-hearted and coerced actions from fully human actions, therefore, is one of saying how it comes about that an agent identifies with her behaviour. Frankfurt's suggestion is that this is accomplished by having an effective second-order volition to be actuated by a first-order desire. The trouble with Frankfurt's account is that although the self-reflective character of second-order volitions captures a sense that free actions must be somehow self-conscious, having effective second-order volitions is not enough to explain how someone's behaviour can be thought of as truly her own doing, because there is nothing about second-order volitions that necessitates an agent's identifying with them: it is just as possible for someone to be alienated from her second-order volitions as from her first-order desires. Frankfurt's initial response to this problem was to say that if an agent is alienated from any higher-order volition of hers, this means simply that she has yet-a-higher-order volition in respect of that volition.[12] Since it is possible to put the question in respect of any logically possible order of volition, the problem with Frankfurt's account has been perceived as one of infinite regress. As Gary Watson has argued, however, the problem is not that the ascent of potential higher-order volitions can never be completed, but that there is nothing about higher-order volitions, as such, that necessitates identifying with them: 'higher order volitions are just, after all, desires'.[13] If higher-order volitions have special significance for the agent, it can only be because she has given them that significance or because they have that significance for her, and this entails that as the one for whom they have that significance, and who may be said to bestow such significance on them, she occupies a position that is logically distinct from them. Frankfurt has sought to overcome this problem by positing a 'decisive act of identification' or 'decisive commitment' in favour of some higher-order desire.[14] But if 'decisive commitments' is taken to refer only to those commitments an agent actively makes, then Frankfurt has begged the question, since his account is supposed to explain how it is that an agent can actively intervene in her behaviour or, in other words, what it is that constitutes her intervening.[15]

Watson argues that what distinguishes those cases in which an agent can be said to identify with her behaviour from compulsive behaviour in which she does not is that the desires and emotions that produce the latter are 'more or less radically independent of the evaluational systems of these agents'.[16] Whether one identifies with one's behaviour, according to

Watson, depends on the degree to which it reflects one's evaluational system, or in other words, on what one takes to be worthwhile, which includes long-term goals as well as normative principles. Desiring something is not the same as judging that it is good or worthwhile. Watson has subsequently rejected his evaluational account of identification, partly on the grounds that it conflates evaluation with judging good, but also because it is possible to value things in particular cases that are not sanctioned by one's 'general evaluational standpoint'.[17]

Watson's evaluational account is also open to the objection that it is just as possible for an agent to be alienated from the values that motivate his behaviour as from his higher-order desires. A person might judge, for example, that the values that motivate him are too materialistic. It will not do, according to Velleman, to argue that in such cases the values the agent is motivated by are no longer integrated with his 'evaluational system', for then some account would have to be given of the difference between values that are integrated within an agent's value system and those which are not. One would naturally think that the distinction is between those values the agent embraces and those he does not, but such an explanation would presuppose, once again, the activity of the agent, which the distinction was supposed to explain.[18]

If identification cannot be guaranteed by higher-order volitions or the agent's evaluational system, what else is left, among the various mental antecedents of actions, that will guarantee it? Since values and higher-order volitions appear to have been ruled out as candidates on the grounds that it is possible for agents to adopt an attitude of critical detachment from them, if there is a mental state or event that can occupy the agent's functional role, other than the agent herself as a ***sui generis*** entity, it must be such that the agent cannot adopt a critical attitude towards it,[19] or if, as I shall argue, it is a sequence of events, one at least of its components must meet this requirement. What can it be?

We seek an analysis of acting decisively in a full and robust sense, which captures the idea of an agent's attitude to her actions (and the motives for them) being such that she identifies herself with them, or, in other words, such that she is not alienated from them. Our belief is that an agent so identifies with a contemplated action when she reflects on opposing reasons for acting, decides in favour of one (and, therefore, against its competitors) and acts accordingly. She fails to identify with an action if the reason for acting in whose favour she judges is not the one that motivates her to act.

Consider the case of an agent reflecting on competing reasons for action. Let us suppose that on weekdays Lucy regularly sets aside the last

two hours before going to bed, for home study towards a long hoped-for qualification. Today, however, she has organised a successful promotion, which has gained her some acclaim but has left her feeling drained. Should she treat herself to two hours of mindless gratification in front of the television or keep to her usual schedule? She may ask herself whether, if she were to opt for the TV, she would just be giving in to her lethargic desire to do nothing. And since gaining her qualification calls in general for her to deny herself such gratification, doesn't that imply that on whatever occasion the question comes up she must deny herself? Alternatively she may think that it is good for people to give themselves occasional rewards for achievement, such as a night off from study, and that always to go in for rigid self-denial can tend to wear down one's energy and enthusiasm. Suppose, however, that it then occurs to her that this belief is merely a rationalisation of her desire to indulge herself. Perhaps she finally settles the question to her satisfaction by noting that she has had that desire several other times, in which there had been no occasion for rewarding herself, and that she did not give in to it. This time, she decides, is different. She takes the night off and feels secure in doing so.

One can imagine situations in which the process of reflection is less straightforward, the eventual decision less conclusive, and one's attitudes toward it less than wholehearted. In those cases it might be impossible to say either that one identifies with one's decision or that one feels alienated from it. But acting and deciding wholeheartedly - fully identifying with what one does - is at least something of which we can conceive.

Among the reasons for acting that were critically assessed by Lucy we find both desires and values. The qualification she seeks is something she values and desires. Her usual rule of self-denial is a standard she aims for on a day-to-day basis, and therefore something she can be said to value. She desires, in addition, to take the evening off. Her practical reflections, therefore, would have involved critically scrutinising the values and desires that would be realised or satisfied by the competing courses of action she was considering. Her critical detachment from these desires and values would have consisted in comparing and evaluating them and in putting certain questions to herself about them. Her critical distance from her desires and values, in other words, would have been constituted by the process of practically reflecting about them.

What, then, did her identification with her reasons for taking the day off consist in? Or, put another way: what would have been required for her to have been alienated from what she eventually did? To be alienated from what one does is to fail to do what one really wants to do. This is a tautology, but worthwhile, nevertheless, in keeping to the fore that what

we are after in seeking to explicate what it is for an agent to identify with what she does is just what it is, if anything, that enables us (and her) to say what it is she really wants. In this case what Lucy really wanted was to take the evening off. She was not sure, initially, that this was what she really wanted. She came to the conclusion that it was what she really wanted by reflecting about it. She arrived at that judgement by asking herself whether satisfying her desire for a relaxing evening was compatible with something else she valued, and which she desired more strongly, and also by putting a question to herself about whether her overriding values about achievement and self-discipline should sometimes be less rigidly adhered to. Her reflections revealed that she had both values and higher-order desires relating to what she should do,[20] but neither her having higher-order desires nor her having values explains how it came to be that what she really wanted was an evening off. Reflection on her values and higher-order desires could conceivably have produced a different judgement about what she really wanted. She might have asked the same set of questions and come to the conclusion that the argument about rigid self-denial was, after all, just a rationalisation. Alternatively, she might have judged that although rigid self-denial can have bad consequences, doing her homework had lately become so burdensome that if she faltered in her routine she would give it up completely. What such a judgement would have revealed to her, and what her actual judgement did reveal, or at any rate what she would have sought from such a judgement, was which action would best realise ***what mattered most to her*** or what she cared most about. And what mattered most (what she really wanted) was not a question of which desire was strongest, or occupied the highest order of preference, or even which of her values were most fundamental, but what was best, taking into account all the considerations (which might have included her goals and ambitions, her standards and self-image, her social context, her needs and desires, and the time and place) she recognised as needing to be taken into account.[21] What she really wanted was what seemed best, taking into account the several and diverse considerations agents employ in deciding such matters: which is to say, what she cared most about; what she valued.

Watson's account of identification as attaching to actions that are motivated by one's system of values, therefore, is (despite his disavowal of it) essentially correct. For a person to determine what she really wants requires that she brings those things to bear that matter to her, in scrutinising, evaluating and judging her competing motives; and 'those things that matter to her' is just another way of describing her values, or what it is that she values.

It might seem as if Lucy failed to act in accordance with what mattered most to her (the sought-for qualification), in failing to do that which would contribute to bringing it about. Her judgement, however, was that attaining the qualification she valued and desired would not be substantially affected by allowing herself immediate gratification on a special occasion, and that adhering too rigidly to her usual standards of self-denial and achievement could be destructive. Her value system, therefore, included normative principles about the way in which her other values should be pursued and realised. When she decided to take a rest, she acted in accordance with her value system, or rather it seemed to her after careful reflection that this was the case. She could of course have been mistaken, and might later have come to such a judgement. In that case, she would have judged herself to have been mistaken about what would best realise what mattered to her. But for as long as it seemed to her that the judgement she made was the right one - that giving the books a miss was the best thing to do, given what she cared about - she could not have failed to identify with her decision.

To be alienated from what one judges best, given what one cares about, would involve a contradiction, since to be alienated from a motive or the action it causes is just to judge that it is not best, ***given what we care about.*** Critics of the evaluational account of identification have argued, as I have said, that it is possible to be alienated from the values that motivate one's behaviour and that it is possible for an agent to identify with actions that are not motivated by her values.[22] The latter claim seems to me to be false, and I will have more to say about it below. The former claim is true, but this does not in any way undermine the proposition that an agent identifies with what she judges will best realise those things that matter ***most*** to her.

In fact there is no other way for an agent to assess and evaluate the competing motives on which she might act, except in relation to those things she values. Reasons for acting are not self-evidently good, bad, right or wrong, but can only be judged as such against some standard that is independent of them. Without a system of values we would have no basis for critically evaluating contending motives: we would simply be carried along by the psychologically strongest contender (relative to what the Will is most conditioned to respond to). Thus, if an agent is alienated from any value that motivates her actions, this can only be because she has other values that are incompatible with it, and which are more strongly held or more fundamental to her. As I mentioned above, a similar defence of Watson's evaluational account of identification is anticipated by Velleman, who argues:

> Of course, Watson refers not just to values lodged in the agent but to the agent's evaluational system; and he might argue that values are no longer integrated into that system once the agent becomes alienated from them. But in that case, Watson would simply be smuggling the concept of identification or association into his distinction between the agent's evaluational system and his other, unsystematized values. And just as Frankfurt faced the question how a volition becomes truly the agent's, Watson would face the question how a value becomes integrated into the agent's evaluational system.[23]

I suppose that integrity of values is something for which reflective people strive, and that they never entirely achieve. If we perceive a lack of coherence among our own or someone else's values, we describe those values as being in conflict. What would lead us, then, to say that among competing values some are more truly our own, whereas others are less so? The question is one that is usually best answered by the individual whose values are in conflict. It might not be something that is immediately apparent to her, and might have to be discerned by sustained reflection. One criterion she might bring to bear in doing this is to ask which values are closest or most fundamental to the sort of person she wants to be or feels herself to be.[24] What she would be attempting to determine by this is not ***per se*** which values are properly integrated with her evaluational system and which are not, but what she values most and what are the practical implications of that. It would be a tortured description to say of a person who recoils from his own materialism[25] that he wishes to downgrade its position, ***qua*** value, in his evaluational system. Rather, we should want to say that he wishes to limit the degree to which materialism motivates his behaviour.

It might be argued that it remains possible, on this account, for an agent to identify the value that occupies the highest position in her value hierarchy, and to turn her critical attention on it. And if 'critical' is to mean anything, that must involve the possibility that she finds it to be deficient in some way, or feels herself alienated from it. It would beg the question to reply to this that she could only critically examine her highest value(s) from the perspective of some still higher value. There is, after all, nothing inconceivable about identifying whatever one values most, putting certain critical questions in respect of it and coming to a negative judgement about it. What is wrong with the objection is the idea of an evaluational system it implies. To have values is not simply a matter of having a list of principles or virtues, of ascending importance, which objects, people, states of affairs and courses of action can be assessed against. Rather,

one's values are embedded and expressed in a variety of practices, behaviours, standards, beliefs, plans, concerns and desires one has about oneself, others (living and dead), society, civilisation, nature, the future, and the other innumerable objects of human interest and concern. If I care a lot about respect and considerate treatment by others, my caring about them does not consist in there being an item with that label in a logbook or a matrix of values, which has a definite weighting (adjustable perhaps), and which my behaviour and that of others can be judged against. My valuing those things consists in the pleasure I may take in the thoughtfulness someone may show me, the degree of distress it may cause me to be treated with indifference, what reasons I take to be sufficient for being inconsiderate to others, and so on. I may not know how much (or how little) these things mean to me until I am confronted with a situation in which their importance is demonstrated with particular forcefulness. That people often do not know what they value until it is denied to them is no less true for being a cliche.

There are in addition a host of rules of thumb, exceptions, modifications and limits, serving a variety of purposes, that we set to the ways we apply our values. To value the splendour and ornate artistry of certain cathedrals and palaces, for example, does not imply that one will want golden chalices, fine tapestries and plate glass windows in one's home, or that one could not value a domicile that is simple and austere. If value systems were like tables of commandments or whatever, we would have to say of those whose taste for finery in public buildings is not matched in the way they decorate their homes that they place a higher value on the principle 'everything in its proper place'. The truth, however, is not that such people care more that everything be in its proper place than they do about the opulence and splendour of cathedrals, but that 'everything in its proper place' would enter into the scope of application of their values.

To judge what we care most about, therefore, is not to appeal to one's highest value, nor indeed does it require that one even has such a thing. What matters most to us in any context can be determined by a range of values, operating within the rules of conduct and moral convictions we adhere to, our life hopes, personal loyalties, enmities and aesthetic preferences. What the objection conceives of, therefore, would be the critical assessment of a value *qua* value, considered in abstraction, rather than of the things we value; which would be appropriate for purposes of conceptual analysis, but not for reflecting on what we care most about. To feel that one is too materialistic is not necessarily to judge that materialism is being given too high a weighting relative to other things one cares about, but that one judges that in certain situations one gives too much weight to

it, which need not imply that in other situations one gives any thought to it at all. Personal values are embedded in one's life and cannot be judged or assessed in abstraction from it. It follows from this not only that we do not apply our values by checking people, things, actions, and so on against value tables, but also that what we value is rather more diverse than philosophical discussion often allows for and may include sometimes doing things that would appear to be at odds with what we most value.

Watson was led to reject his earlier values-based account of identification partly because it is possible to value things in particular cases which are not sanctioned by one's 'general evaluational standpoint'.[26] This may be the case if, to take an example from Watson, one does something because it is thrilling or exciting, although at another level one judges it to be vulgar and demeaning, 'but when it comes down to it, one is not (as) interested in that'.[27] The objection, however, can easily be accommodated by the evaluational account of identification. Notice that in this example the agent's reasons for giving herself up to excitement and thrills include a second-order judgement about the relative importance, on this occasion, of the further judgement that the activity may be vulgar or demeaning. To judge in this way a person does not simply give herself up to the moment, but does so for reasons that have application beyond it. A judgement that on occasion thrills and excitement should take precedence over dignity and good taste would imply that it is sometimes worthwhile to seek thrills and excitement, perhaps because it is healthy and human to do so. In that case, such a judgement would, contrary to Watson's claims, express the agent's values (rather than her 'general evaluational standpoint'). A judgement of this sort could cohere with a set of values that might ordinarily rank certain activities as vulgar or demeaning, by placing a limit on the application of such values in the interests of health or balance. For sure, an agent who acts in this way might be kidding herself in thinking that her actions are in keeping with her values, and might subsequently regret giving in to her desire for thrills.[28] In that case, however, she would also become alienated from her frivolous actions (after the event). But for as long as she takes the view (even if mistaken), that going for thrills is what matters most at the time, she will understand herself to be acting in accordance with her values and will identify with her actions.

If instead someone consistently gives herself up to thrills and excitement whenever the occasion presents itself, she could not realistically be understood as thinking of such activities as demeaning. Or again, if, contrary to Watson's example, she does not consider what she does to be vulgar, then there would be nothing in her values for what she is doing to run counter to. Or, finally, if her values are such that she would ordinarily

regard such activities as vulgar and demeaning, but she forgets about that on this occasion and is carried away by the moment, then what she does is not something with which she could be regarded as fully identifying.

If an agent embraces an action without considering it best,[29] all that implies is that the word ***best*** is being used to indicate what is morally best, or what would usually be best, which is entirely compatible with a judgement on a specific occasion that frivolity and excitement are best. Watson's denial of this suggests that he may implicitly be operating with a monolithic conception of an agent's values as some sort of fixed system of rules and principles of the kind I have been criticising, or that he identifies 'judging good' or 'judging best' too closely with moral judgements.

Again, it might be argued that when an agent identifies with a motive, it is too elaborate and contrived to say that this is always preceded by a desire to do what matters most to her, which causes her to reflect on her motives. One can often choose what to do with scarcely a thought about it, and still fully identify with one's choice. Wanting to do what one cares most about, however, does not require that one desires to realise one's most deeply held values, or that one must always reflect carefully and exhaustively on what actions will best realise them. A person's values, as I have said, include beliefs, attitudes, and ideas about their scope and application, and about what situations call for reflection and assessment. Wanting to do what matters most may often mean wanting to do what is right for a particular situation, where that is informed by a sense that it is sometimes beneficial to relax and let go. One needn't go through a process of reasoning through contending motives and values in such a situation, if one has a general sense of its being the kind of situation in which one can properly allow oneself latitude in relation to certain values and in which it is inappropriate to 'think too much' about what to do. It is not that agents reject (temporarily) their deeply held values in such situations, or arbitrarily set them aside, but rather that they also value sometimes being able to satisfy desires that do not accord with some of their deeply held values. Desiring to do what matters most, therefore, can sometimes mean just wanting what is right for the moment.

The reader who remains unpersuaded by this might reflect on the kind of example of temporary value rejection provided by Watson, in which an agent opts to do what she otherwise considers to be vulgar and demeaning. Would the example seem so persuasive if the agent were to choose instead to do what she believes to be corrupt and depraved, or what she knows to be vicious and cruel? We know that people who disvalue viciousness and cruelty do not take time out from ***those*** values because 'when it comes down to it, [they are] not (as) interested in that'.[30] One who 'on occasion'

identifies with and acts on desires to be vicious and cruel could not properly be understood as disvaluing such behaviour. What is the difference between what we would say about giving in to viciousness and about giving in to vulgarity? Just that a shared sense of the scope of values about doing what is vulgar and demeaning contains enough slack for us to believe that it is sometimes healthy to ignore them, whereas our shared sense of the scope of opposition to viciousness and cruelty is not such that it is acceptable on any occasion to set it aside.

It might be maintained, I suppose, that to claim that someone might, on occasion, value what she otherwise considers to be vulgar and demeaning is incoherent, since to consider a way of acting to be demeaning *just is* to disvalue it. This again would be open to the objection that it treats personal values as monolithic systems or rule books, which does not allow for restrictions on the scope of values and for valuing occasional exceptions to acting in accordance with one's values. But even if this view of value hierarchy were true, which it is not, the objection would still be guilty of mistakenly equating the problem of whether an agent is correct in determining what matters to her with that of identification. If an agent believes that x-ing is what she most values on some occasion, and she is mistaken about that, it will still be the case that she identifies with x-ing because and for as long she believes that it is what matters most. The evaluational account does not claim that identification consists in acting on what one cares most about, but rather that it consists in making a judgement about what one most cares about.

So long as an agent has satisfied herself, therefore, on the basis of whatever reflection on her contending motives she takes to be sufficient, that a course of action is the one that will best realise what matters most to her, she identifies with that motive. Velleman might object to this that the concept of identification is still being presupposed, rather than explicated, via the notion of what we care most about, and that it is the notion of what we care most about, or what matters most, that now needs explaining. Certainly, if what we care most about, or what we value most, were settled by some radical act of choice, we would be no further forward in explaining how agents identify with what they do, since we would now have to explain how it is that an agent can make a choice in such a way that she cannot be alienated from it.[31] But determining what one cares most about is a discovery rather than an act, and to identify with an action is likewise primarily a matter of discovery, realisation or coming to believe that it will best realise what we most want. Identification is not an act: it is a discovery or realisation.[32]

If we say that identification is a discovery rather than an act, however,

although we can no longer be accused of presupposing what our account purports to explain, or of leaving it a mystery, we now face the problem of saying how it is that identification has anything to do with an agent's ***participation*** or ***intervention*** in her behaviour. If identifying with the motives that actuate one's behaviour is simply a matter of ***realising*** (or coming to believe) that they are in line with what one care's most about, identification seems merely to be a matter of observing, from a non-participatory point of view, that one's actions and motives are what one would wish them to be. If identification is necessary to our notion of full-blooded agency, therefore, it is not sufficient for it.

According to the passage from Thomas Reid I quoted earlier, the position of the agent in relation to his conflicting motives is like that of a judge hearing the pleas of contending advocates, which implies not only that the agent passes judgement on his motives, but that the judgement determines the outcome of the proceedings. Others, as I mentioned earlier, have spoken of the agent throwing his weight behind the motive with which he identifies.[33] If an agent does not throw his weight behind a motive merely by identifying with it, how does he do it? He does so by discovering which motive best realises what matters most to him, and thereafter by strengthening that motive with the force of his desire to do what matters most to him. If it is true, as it seems to be, that we are sometimes able to choose to do what we desire less, what enables us to do so is that we are led, by a desire to do what we care most about, to critically assess the contending motives; that we judge that the weaker motive is more in keeping with what we care most about (thereby identifying with it); and that this motive is strengthened by the independent force of our desire to do what we care most about. And likewise, if the motive with which one identifies is not the one that actuates one's behaviour, this happens because the additional force of the desire to do what matters most is insufficient to prevent the opposing motive from triggering one's will.

REASON, VALUES AND DESIRES

It is therefore the desire to do what we care most about - in other words, to do what we really want - which initiates the process of practical reasoning that leads to identification with a reason for acting, and which lends additional causal weight to that reason. This is in line with David Velleman's argument that the mental event or state that plays the functional role of the agent, and with which she cannot fail to identify, is a 'motive that drives practical thought itself'.[34] Velleman misidentifies this, however, as a

'desire to act in accordance with reasons'. Assuming that we ever act on such a desire, it cannot be this that initiates the process of our coming to identify with a motive for acting, since such a desire would be realised by motives and actions from which we can be alienated: acting in accordance with such motives would still be acting in accordance with reasons and would therefore satisfy a desire to act in accordance with reasons. Velleman suggests that when a desire on which we might act

> appears to provide the strongest reason for acting, then the desire to act in accordance with reasons becomes a motive to act on that desire, and the desire's motivational influence is consequently reinforced.[35]

In that case what Velleman has in mind is a desire to act in accordance with one's strongest reasons; and we might judge this to be close enough to a desire to do what we care most about, or what best realises what we care most about, to make quibbling over the precise wording seem pedantic. What this would overlook, however, is that having rejected the agent's values as the source of identification, Velleman has no account of how some reasons for acting can be stronger than others. Clearly the relative strengths of reasons for acting must depend on who an agent is, on what her circumstances are, and, most importantly, on what she cares about. Reasons for acting are not intrinsically strong or weak, but are only so in relation to what we want or what we care about.[36]

Velleman has indicated (in correspondence) that the 'desire to act in accordance with reasons' has a schematic role in his account of full-blooded agency, in which the reference to reasons could be replaced by whatever reasons are found to consist in according to a correct substantive theory of the content of reasons for acting. Such a theory might show that reasons do consist in values, and nothing Velleman says is meant to rule out such a possibility. My objections to this are first that a desire to act in accordance with one's best reasons fails to express - even schematically - the source of motivation in all conceivable examples of full-blooded agency. Velleman suggests that although a desire to act in accordance with reasons may not form the ***de dicto*** content of one's reasons for acting, one could still be said to be motivated by it, if one were acting on

> a desire to act in accordance with considerations of some particular kind, which happened to be the kind of consideration that constituted a reason for acting ... [such as] a desire to do what makes sense, or what's intelligible ... in the sense that [one] could explain it.[37]

But it is possible to think of examples of acting in the full-blooded and distinctively human manner we have been considering which cannot be construed or paraphrased in this way. A ladder collapses, leaving a house-painter hanging - a la Harold Lloyd - from the guttering of a three-storey house. She has two options: either to remain clinging to the guttering or to let go with one hand and attempt to swing and catch at the branch of an overhanging tree. The thought of swinging for the branch is terrifying, but she realises that if she remains where she is it is unlikely either that the guttering will support her weight or that she will have the strength to hold on long enough for help to arrive. She decides that her only real chance is to swing for it, and, facing down her terror, she lets go, swings, catches the branch and climbs down the tree. What makes this a case of fully human agency is the agent's capacity to critically assess her options and to come to a judgement that enables her to overcome her fear of letting go. What motivates her actions is a desire to survive, and although that desire can be schematically expressed as wanting to do what matters most, it cannot properly be described as a desire to act in accordance with reasons such that she desires to do what makes sense to her or what she could explain.

Even in cases of full-blooded agency where we might think of ourselves as wanting to act in accordance with certain reasons - in which we are concerned that our actions meet certain standards - it is not a concern with reasons that is the fundamental source of our motivation, but desires to do what will be most fulfilling, to be better people, to act honourably, or to be true to ourselves: desires that are all to do with what we care most about.

Velleman is also mistaken in thinking that it is only the desire that drives practical thought that occupies the agent's functional role. Where practical thought is motivated by the desire to do what matters most, it is not only the desire that fulfils the agent's role but also the process of practical reasoning itself. Practical reflection can make a difference to what we do in a way that does not derive entirely from what initiates it (the desire to do what we care most about). One can reason well or sloppily, and whichever is true can make a difference to what one perceives as the course of action that will best realise what one cares most about. That one identifies with a course of action, therefore, is secured by the desire to do what matters most causing critical assessment of contending motives; but *which* course of action one identifies with depends on the process of critical assessment itself, rather than just what causes it. That the process of practical reflection is integral to what we take ourselves to contribute to our actions is borne out by the fact that in what we take to

be the full-blooded exercise of agency it is the agent that does the weighing up and it is done actively. As it stands, Velleman's account of agency is compatible with our being instructed by some external agency, or by inner voices, to believe that some reason is the strongest (i.e., if the role of agent is discharged by a desire alone).[38] The agent must not only throw her weight behind motives, but also assess which motives to throw her weight behind.

One obvious objection to the claim that practical reason, as such, also forms part of the agent's functional role, is that, unlike the desire to do what matters most, it is possible to be alienated from practical reasoning. Indeed this has already been given as an argument against Velleman's 'desire to act in accordance with reasons'. If someone is figuring out the best way of doing something she doesn't really want to do, she will reason half-heartedly. What follows from this is not that practical reason cannot be regarded as playing the agent's role, but rather that practical reason that is not motivated by the desire to do what matters most cannot do so. And to be alienated from practical reasoning that is directed at what one does not really want requires that one also reasons about that, and that doing so is motivated by the desire to do what matters most. To play its part in the full-blooded exercise of agency, therefore, practical reasoning must be appropriately motivated. This does not entail, however, that the part it plays is entirely reducible to what motivates it, since, as I have argued already, *how* one reasons can make a difference to what motive one comes to identify with, and therefore to what action one performs.

A further objection that might be made to the claim that practical reasoning is an essential constituent of the agent's functional role in full-blooded agency is that, like Frankfurt's account of decisive acts of identification with higher-order desires, it presupposes the concept it is supposed to explain. The various components of practical reasoning, such as questioning, comparing, analysing and evaluating, are all properly described as actions or activities. Their being mental does not disqualify them from counting as actions, or certainly not in any sense that interests us. The account need not be circular, however, if the category of action it presupposes is that of basic rather than full-blooded agency. And in themselves, the various actions that are involved in practical reasoning are exactly that: basic (mental) actions, rather than actions of the sort that satisfy our full-blooded account of agency. Therefore, although, as I suggested earlier, what distinguishes basic actions from involuntary behaviour is a

separate problem from what distinguishes full-blooded agency from half-hearted or addictive behaviour, an adequate account of basic actions is a precondition for an account of full-blooded agency.

According to our causal analysis of basic actions, someone engages in practical reasoning

> *if* she wants ϕ and believes practical reasoning is a means to bringing about ϕ and that there is no better alternative means to bringing about ϕ, and knows how to reason practically, and is able to reason practically, and has inferred[39] from this that practical reasoning is worth doing, and has a will that has been conditioned or habituated so as to be triggered by wanting and believing all of this.

If, moreover, what the agent wants is to do what matters most to her, and her basic acts of practical reasoning are motivated by that, then they constitute one element of her functional role in relation to actions in which she can be said to participate fully.

Our understanding of practical reasoning as something an agent does is not such as to require that the agent intervenes in it, or even that she critically assesses her motives for doing it. She can do both,[40] but her reasonings will not fail to qualify as actions in the basic sense as a result of there being no critical intervention in them, and nor will they be disqualified from fulfilling her functional role in any more complex and full-blooded action, just as long as they are motivated by the desire to do what matters most.

The notion of critically assessing one's motives is one of an intentional activity, and therefore calls for an account of action that is prior to the overall account of full-blooded actions to which such critical assessment is thought to contribute.

FREE AGENCY

An agent acts in the full-blown distinctively human sense, therefore, whenever she identifies with a motive and that identification leads the motive to be strengthened, causing her to be actuated by it. This happens when an agent's desire to do what matters most causes her to critically assess her contending motives, which leads her to judge that one motive best realises what matters most to her, and the additional force of her desire to do what matters most is thereby added to that motive, thus causing it to be the one that triggers her will to action.

As I have argued, identification, even if arrived at by appropriately motivated practical reflection, is not enough for full-blooded agency. Our conception of fully human agency will not be satisfied if it happens that an agent identifies with the motive that triggers her will, but would have been actuated by that motive even if she had been alienated from it. One who embraces her heroin addiction must also embrace a certain kind of slavery. When an agent has conflicting motives, to act in the full-blooded sense she must not only add her weight (the weight of her desire to do what matters most to her) to the motive with which she identifies, by identifying with it, but her doing so must be necessary for that motive to be the one that actuates her.

Is our causal analysis of full-blooded agency enough to identify those actions that are free in the sense that they realise what the agent most wants and are performed for that reason? It is more than enough. Our conception of full-blooded agency is one in which the agent intervenes in her behaviour by critically assessing her contending motives for action and throwing her weight behind the motive with which she identifies. It would be perverse to say that people act freely only when they have conflicting motives and can identify with one or the other. On the other hand, neither would we want to say that an agent is always free whenever she does not have conflicting motives. Where is the line to be drawn? The answer, again, is that a person cannot be said to have acted freely if, were she to have judged differently, she would have acted on the same motive. What this means in terms of our account of full-blooded agency is that a person acts freely if, when she has conflicting motives, she can critically assess those motives and be actuated by whichever motive she identifies with, or, where a person does not have conflicting motives, that were she to have had reasons for acting differently, she would have been able to critically assess her conflicting motives and would have been activated by whatever motive she identified with.

Libertarians may object at this point that this kind of counterfactual responsiveness to identification with motives is insufficient for free action since it requires only that the agent can throw her weight behind whatever motive she identifies with, which is compatible with her being unable not to identify with it. Since the objection is about being able to identify differently than one does, it is different from the claim that free will requires being able to act otherwise than in accordance with what one really wants/what is right and reasonable, which I rejected in Chapter 3. How can the objection be met or dispelled?

An agent who reflects on conflicting motives may believe that given her values, and the situation, no other conclusion is possible than the one she

arrives at. She may still be able to judge, nevertheless, that she could have identified differently, were she to have had reasons for doing so, provided that at no stage before or during her reflections was there anything external to those reflections necessitating her to identify as she did, and that she experienced nothing as necessitating her to reason as she did other than it seeming to her that this was the optimum way of coming to a judgement about what mattered most to her.

In Chapter 4 I argued that even when there is a high expectation, on inductive grounds, that one will end up deciding in a particular way, there is never complete certainty about this before one has finished deliberating, and where there is complete certainty one cannot understand oneself as genuinely deliberating about what to do. Although, as I have argued, identifying with a motive, unlike making a decision, is not an action, the same argument applies, since both are arrived at by deliberation (and often enough will result simultaneously from the same process of deliberation):[41] for as long as an agent believes deliberation to be necessary she cannot be absolutely certain about its outcome. Since the only thing that could guarantee certainty about the outcome of an unresolved decision would be a valid prediction based on a true and precise identification of an occurrent or antecedent mental state as being of a particular causal type, and since no such identification can be true for the agent who must make the decision, there cannot be certainty about one's own future decisions.

The same goes for identification. For as long as an agent perceives practical reasoning to be necessary, she cannot be certain about what motive or action she will identify with, and it must be possible for her (even if it seems very unlikely) that she might identify with the opposing motive to the one she expects to identify with. And as I have argued in respect of those difficult decisions that involve real deliberation, the possibility that she could identify differently than she does is no mere epistemic possibility - that she might identify with either motive 'for all she knows' - but, since no prediction about the way she will identify can be true for her, it is literally/nomically possible for her that she might do so.

Therefore, if inductive evidence cannot cause an agent to experience identification as necessitated, she will not experience anything other than her own reflections as leading her to identify as she does, and for as long as she continues to reflect, it will be open to her (even if it seems unlikely) to identify with any of the contending motives she is considering. This also fits with situations of strongly contending motives being our paradigm case for full-blooded agency. The more sharply an agent's motives

conflict, the less certainty she will have about how she will identify and what she will decide to do, and the stronger will be her sense that it is up to her what to do.

What else could a libertarian ask for? It would be strongly counterintuitive to suggest that free agency requires that we be able always to identify differently, or to experience ourselves as able to do so, such that if we know what we really want and have no doubts about it, we could still have it or feel it open to us that we could identify with an opposing motive or course of action. It would require, for example, that should an agent identify with a course of action she believes to be conducive to her own well-being or that of someone she cares for, she should have it open to her that she could really want to do what would be ruinous to that well-being. Who but a psychotic could ever have such a capacity, or feel herself to have it, or consider herself to be unfree in lacking it?

There is one remaining difficulty with the account of free agency to which we have come. I have argued that self-determination requires that agents are sometimes able to select their own reasons for acting and to intervene decisively in the initiation of their own actions. An agent judges between conflicting reasons for action whenever practical reasoning, motivated by the desire to do what matters most to her, leads her to identify with whatever reason or course of action it appears will best realise what she cares most about. An agent intervenes in her behaviour whenever it is the case that the additional force of her desire to do what matters most to her is necessary and sufficient for her to act or decide to act in the way with which she identifies, and that the action acquires that force as a result of her identifying with it. If it happens instead that following identification with a motive an agent is actuated by an opposing motive, her action will not be self-determined and she will not regard it as such. That she does not regard her action as self-determined, of course, is simply another way of saying that she is alienated from it.

According to the hypothesis about the Will, an agent will not do what she really wants because it is what she really wants if her will has been so conditioned in respect of certain motives and situations that it is not responsive to what she really wants; although prior conditioning of her will can also protect her from giving in to strong desires with which she does not identify. However, if whether the Will is triggered by what the agent really wants is not itself under the agent's control, we would have to think of what results from it as not being really under her control

either.[42] As it stands, therefore, the hypothesis about the Will seems open to the objection that it rules out self-determination not only for actions where the Will is not triggered by what the agent most wants, but for all actions. The answer to this is that according to the hypothesis, the Will is open to modification or conditioning. If one's will has been conditioned to be triggered by desires from which one is alienated - a craving for cigarettes, say - one can attempt to overcome this by 'modifying strategies'. Such strategies would include engaging in mental routines, like thinking about why one does not really want to satisfy such desires, how it would be and how one would feel if one were to act in accordance with what one really wants, seeking help and therapy from others, placing oneself in environments and situations in which one's unwanted desires will be less likely to be prompted, acquiring habits that tend to diminish the force of the unwanted desires, breaking habits of thought, and so on.

The account of self-determination I argued for in Chapter 3 is now complete. We now know what it means to say that someone can will in accordance with what she really wants, and how it is possible to acquire that ability when it is lacking.

This account of self-determination, moreover, accords with how we would prereflectively distinguish between the presence and absence of self-determination. Alan wants very much to bring his cigarette addiction to an end but is seemingly unable to do so, since despite what he wants he persists in smoking and his attempts to give it up have repeatedly failed. According to our account, his habitual smoking has conditioned his will to respond to his craving, so that it is unresponsive to what he really wants. Say then that he finally succeeds in overcoming his addiction by enrolling on a 'make or break' course in which he is given nausea-inducing anti-nicotine drugs, counselling about how to avoid being led back to smoking, and an environment that isolates him from exposure to temptation. The effect of the course is to condition his will so that it will not be triggered by his desires for cigarettes, and will respond instead to what he really wants. If he successfully overcomes his addiction, he ends the determination of his behaviour by his past habit. He will have increased his self-determination by causing his actions to be motivated by what he really wants.[43]

CONCLUSION

My actions make a difference to how things would otherwise be in the world, and the difference they make traces back to me as their author. Nor

is that all. I can rightly think of my actions as my own, in the fullest sense; or if I cannot, I have some idea about how to bring it about that they are. **If** my actions (sometimes) match up to the account of full-blooded agency set out above, I freely act in a way that addicts, children and animals do not. So it will seem, and so it will be.

Individuals sometimes imagine themselves to be acting as free agents when they are not. They are not because their behaviour is produced by deranged fantasies and grotesque suspicion. Many of us will sometimes act and behave unfreely, but will be prevented from realising it because we rationalise and attribute false motives to ourselves. It is possible, therefore, to have illusions about free agency. But there is nothing inherently or necessarily illusory about it. Moreover, whatever startling indeterministic tendencies there may be in our world, our conception of free agency can be realised in a universe that is causally necessitated down to its last detail. **If** that conception is an illusion, it will not be on account of the causal necessitation of actions. It would be illusory if human behaviour were causally undetermined in any of the various ways libertarians have claimed that it is.[44] The only good reason for thinking that actions are undetermined in the way libertarians have claimed, however, is that this was thought to be necessary for free will and agency. But it is not. We have good reason to believe, therefore, that we do act in the way we experience and understand ourselves as doing, and that we sometimes do so freely.

7 Conclusion

In Chapter 1 I argued that the traditional argument about the compatibility of determinism with free will and moral responsibility has resulted in a stalemate, in which, although representatives of the opposing camps continue to present sophisticated arguments for their respective positions, there is little prospect of let-up in the entrenched character of the controversy. Some have argued that this can be explained by the existence of contradictory metaphysical attitudes we all share about the initiation of actions, while others have argued that terms such as ***free*** and ***responsible*** lack any real meaning and simply reflect radically divergent and unregimentable images and intuitions. If either explanation should have proved correct, it would have followed that various beliefs we all share about our abilities as agents to freely act and decide are founded (in whole or in part) on illusion. As I hope I have shown, while we can sometimes be mistaken in thinking that we have acted freely, there is nothing inherently illusory in our beliefs about free will and agency (or at any rate, not if determinism is true). Not all of my arguments have been about the experience of acting and deciding, but the idea that the solution to many of the difficulties about free will is to be found in an analysis of the phenomenology of freedom and agency has been a guiding thread.

The principal reason for the inability to make any headway in one of philosophy's most entrenched arguments has been a narrow preoccupation with the meanings of the key terms: a preoccupation that has been reinforced by the twentieth-century 'linguistic turn' in philosophy. If the focus on meaning is supplemented (rather than replaced) by a more careful attention to the experiences that shape our understanding of free will and agency (and I do not imagine that this has been exhausted by what I have said) there is a real prospect of moving beyond the unyielding nature of the traditional argument.

Since the main aim of this book has been to show that incompatibilist worries about determinism are unfounded, and that their desiderata can be met without recourse to indeterminist theories of free will, it will be useful to adumbrate the main conclusions that have been reached:

(1) That we do have sufficient reasons, when confronted with wrongdoing, for blaming, punishing and resenting, even though there are always reasons, in any particular case, for withholding the reactive attitudes and practices.

(2) That someone has free will, or that she is self-determining, if she can do what she really wants because it is what she really wants.

(3) That there is nothing more we could want by way of self-determination than what is specified by (2): the experiences of ultimacy and of being able categorically to choose either way are features of our sense of agency rather than of free will.

(4) That if we have self-determination as specified in (2), and determinism is true, we are not obliged by the fact that our actions are necessitated (in a special sense) to consider ourselves in the grip of the past.

(5) That our experiences, while deliberating, of being able to decide either way are not illusory: it is a logical consequence of our position as deciders that we can categorically decide either way, and not just that we can do so 'for all we know'. No thought about what someone else could in principle know about us, or about what we might know in the future, can give us cause to doubt this.

(6) That we therefore have what incompatibilists have described as ***ultimacy***. When we are faced with difficult decisions it is entirely up to us how to decide. Determinism cannot relieve us of that.

(7) That inasmuch as our futures depend on our decisions, especially our difficult decisions, our futures are open. To the extent that one's future depends on what one decides, the truth of any statement about it is indeterminate.

(8) That our experiences of basic actions are veridical: we do act just as we understand and experience ourselves to do.

(9) That we sometimes have dual-control and dual-rationality, i.e. not only can we sometimes categorically decide either way, but we can sometimes rightly consider that we will be in control of what we do and that we will be acting rationally, however we decide.

(10) That we sometimes have the capacity to act and decide in a fully human sense, i.e. that we have a capacity to weigh up our reasons for acting, to identify with some reasons against others and to act accordingly.

(11) Therefore, that we can be self-determining in the sense specified in (2).

Can incompatibilists still maintain that (1)-(11), or the arguments by which they were reached, do not give them what they want? Of course they can, and as I said in Chapter 1: no doubt many of them will. There is nothing to stop incompatibilists from making their minds up in advance that any compatibilist account of free will must be false, or that only indeterminism will do. But if, as I have shown, the experiences, intuitions and hopes to which indeterminist theories of mind and action have attempted to give expression can be accommodated without recourse to indeterminism, then incompatibilists who insist on exactly what they say they want will be clinging to the lifeless letter of their theories, rather than to what has made the theories attractive or persuasive.

Notes

Notes to the Preface

1. To my knowledge the only contemporary book on free will in which there is any substantial consideration of the phenomenology of freedom is Galen Strawson's *Freedom and Belief* (1986).
2. P. F. Strawson (1982).
3. See Honderich (1988).
4. See T. O'Connor (1993).

Notes to Chapter 1: Are the Problems of Free Will Resolvable?

1. Hume (1975a, p. 95).
2. Kant (1956, pp. 99-101).
3. Davidson (1980c, p. 63).
4. Anscombe (1976, p. 159).
5.

> Liberty, and necessity are consistent: as in the water, that hath not only liberty, but a necessity of descending by the channel; so likewise in the actions which men voluntarily do: which, because they proceed from their will, proceed from liberty; and yet, because every act of man's will, and every desire, and inclination proceedeth from some cause, and that from another cause, in a continual chain ... proceed from necessity. (Thomas Hobbes, 'Liberty and Necessity Consistent', 1962, p. 160)
>
> Is not this a ridiculous liberty? ... such a liberty as a river hath to descend down the channel. ... Such is T.H.'s liberty. (J. Bramhall, 'A Defence of True Liberty', passage quoted in Honderich, 1988, p. 458)

6. P. F. Strawson (1982).
7. Frankfurt (1971).
8. For an account of this kind of philosophical hostility, see Double (1991, p. 6).
9. See, for instance, Adler (1958).
10. Double (1991); Honderich (1988). See also Nagel (1986) and G. Strawson (1986).
11. Honderich (1988, p. 475).
12. Ibid. (pp. 379-450). Honderich argues that questions about the consequences of determinism for what matters to us can be brought under seven headings, concerning our futures, our feelings of resentment and appreciation towards those whose actions affect us, our claims to knowledge, the practice of holding ourselves and others morally responsible, the moral standing or worth of persons, our judgements about the rightness and wrongness of actions, and practices and institutions such as punishment and the state. Each of the questions can bring into view a pair of attitudes - consisting,

for example, in two different kinds of hope about the future and two different kinds of moral disapproval - which produce corresponding responses of dismay and intransigence towards determinism. Cf. Nathan (1992) who suggests that our worries about free will and responsibility result from various possible conflicts between what we want to be the case and what we have reason to believe.

13. G. Strawson (1986, pp. 105-17, 307-17). On the other hand, according to Strawson, the sense of detachment from certain desires, and the concomitant idea of ourselves as free to act independently of such desires, is also a source of our natural incompatibilism (ibid., p. 116).
14. Strawson's account differs from Honderich's, in holding that the 'fundamental sense of the word "free"' (that which 'is in question when it is said that because people are free agents, they can properly be held to be truly responsible for their actions in such a way as to be truly deserving of (moral) praise and blame for them') is the incompatibilist sense.
15. Dennett (1984).
16. Honderich (1988, pp. 475-87).
17. Double's paradigm of the 'pure rational ego' might be particularly appealing to philosophers of a calm, solitary and dispassionate cast of mind, as a model for freedom (Double, 1991, pp. 117-18). Honderich (1988, p. 476) considers the possibility that a philosopher as resolutely compatibilist as Moritz Schlick might unusually lack the attitudes that incompatibilism gives expression to. Galen Strawson also suggests that some of us 'are far more naturally compatibilist than others' (1986, p. 311).
18. Although there are exceptions: Peter Strawson being a notable compatibilist example, and Robert Kane, who has sought to frame an indeterministic account of free will that is consistent with scientific explanation.
19. P. F. Strawson (1982, p. 79).
20. Honderich (1988, p. 492).
21. Double (1991, pp. 6, 114-32).
22. See Honderich (1988).
23. G. Strawson (1986, pp. 28-9). Cf. Honderich (1988, pp. 178-9), and Mele (1995, pp. 221-30).
24. Dennett (1981a); Ginet (1990); Kane (1989 and 1994); Nozick (1981).
25. Honderich (1988, pp. 184-208); Double (1991, pp. 190-211, and 1993). See also T. O'Connor (1993).
26. Honderich (1988, p. 493).
27. Dennett (1984, p. 51); Honderich (1988, p. 471).
28. See Honderich (1988, pp. 400-9), for several examples of situations that can, depending on how one thinks of them, trigger the responses to determinism either of 'dismay' or 'intransigence'.
29. Kapitan (1994, p. 93).
30. This is challenged by Kane, who argues, in chapter 6 of his forthcoming *The Significance of Free Will* (Cambridge Studies in Philosophy), that more needs to be done to trace the roots of our conflicting attitudes about free will.
31. Moore (1912, p. 211).
32. Austin (1961a, p. 111).
33. Chisholm (1964, p. 25).
34. Davidson (1980c, pp. 68-9).

35. 'It is logically possible that as a result of my not willing, not choosing, or not undertaking some action, I might lose any of my powers.... Suppose that I am offered a bowl of candy and in the bowl are small round red sugar balls. I do not take one of the balls because I have a pathological aversion to such candy.... It is logically consistent to suppose that if I had chosen to take the red sugar ball, I would have taken one, but, not so choosing, I am utterly unable to touch one.... I could do it only if I chose to, and I do not' (Lehrer, 1982, p. 44).
36. Davidson (1980c, p. 70n). See also Watson (1987a, p. 158). For further discussion ofLehrer's arguments, respectively for and against, see Anscombe (1976) and Aune (1982).
37. We do sometimes speak of actions as indirectly caused by events or states. To take an example from Richard Taylor (1966), if a person's fear of heights causes him to grip the chair of the ski-lift he is riding, his fear indirectly causes the gripping of the chair by his hands. It does so, however, only by causing *him* so to act, rather than by directly triggering the muscles of his hands (p. 104).
38. It can be argued that although statements of actions and ability do not imply causal conditionals, they do entail them, since the idea that agents initiate actions and have control over them cannot be explicated without the concept of causation. As will be seen in Chapter 5, I am inclined to agree, but entailment is insufficient for equivalence of meaning.
39. Davidson (1980c, p. 76).
40. The label was coined by van Inwagen (1983, p. 16).
41. Cf. van Inwagen (1982, pp. 52-3, and 1983, p. 70). See also Ginet (1966); Lamb (1977); Wiggins (1973).
42. Slote (1982); Wiggins (1973).
43. The argument also avoids the objection of regularity theorists like A. J. Ayer that determinism does not imply necessitation of effects by causes, but only factual correlation of events (1982, p. 22).
44. Lewis (1981).
45. Ayer takes the same view about the status of statements about the past in relation to counterfactual ability claims (1963, p. 239). Cf. Goldstick (1979).
46. Lehrer rejects the strong claim on the grounds that it would require the truth of a conditional analysis claiming that difference in laws or history would follow from the agent preferring such differences. Lewis argues that such a conditional could be true: 'Suppose the agent is predetermined to prefer that there be no difference; had he preferred otherwise, there would have been a difference. (Had anything been otherwise than it was predetermined to be, there would have been a difference in either laws or history)' (Lewis, 1981, p. 116n).
47. Slote (1982).
48. A possible objection to this is that someone might foresee a road accident - because her vantage point enables her to see a van that is about to take a comer that conceals a car reversing out of a drive - without being inclined to say that what was foreseen was not an accident. However, the sense of *accident* that refers to events involving damage to life, limb or property is always relative to what is known or could be foreseen by those involved.
49. Ibid., p. 19.
50. In correspondence, David Cockburn has suggested that statements to the

effect that nothing can change a certain thing that is going to happen in the future - e.g., 'he will be executed tomorrow (so you might as well give up sending telegrams to the President)' - bear out Slote's claim that judgements of necessity are selective. The comparison is unsound. In the first place, Slote's arguments were intended to show that where a statement is entailed or implied by two or more necessary premises it is not always the case that the entailment or implication carries the necessity with it. And whereas it might be the case that some statement about the past, together with the laws of nature, entails my doing something now, a statement to the effect that a man will be executed tomorrow does not entail or imply my doing anything now. I do not deny that statements of necessity are relative or selective (e.g., the statement that nothing can change the outcome of World War II is relative to what might happen now or in the future, and does not entail that Allied victory was necessary before it happened). What I do deny is first that statements about the unchangeability of the past are relative solely to current beliefs, desires, thoughts, etc. (rather than *any* current or future event or state), and secondly that they are implicitly about explanations. Suppose that it (is conceivable that it) should turn out that our current and future desires can affect the past. In that case our current and future desires could be included in explanations of the past, and the past would no longer be unchangeable. Would we say that the past is changeable because our current and future desires can explain it, or because they can affect it? If it is the case that whether a putative explanation really is an explanation (rather than a mere hypothesis) depends on whether it is true, then an explanation of past states of affairs that includes current or future desires would only count as an explanation if our current and future desires really did affect the past. Slote's argument reverses the proper relationship between fact or truth and explanation.

51. Slote (1982, p. 20). See Fischer (1994, pp. 29-44) for another argument against Slote.
52. Cf. van Inwagen (1983, pp. 120-6).
53. See van Inwagen's objections to Flew's 'paradigm case argument' (van Inwagen, 1983, p. 110).
54. Honderich (1988, p. 475).
55. This view is shared by the incompatibilist Robert Kane (1994), although we differ about what those sources are.
56. P. F. Strawson (1982).
57. Ibid., p. 66.
58. For interesting and useful discussions of Strawson's distinction between the objective and reactive attitudes, and the criteria for adopting them, see J. Bennett (1980) and Watson (1987b).
59. P. F. Strawson (1982, p. 70).
60. Double, for example, puts the objection that even if Strawson's account has captured the logic of the reactive vocabulary, his account still falls short of justifying resentment and moral responsibility (1991, p. 83). Susan Wolf, likewise, takes Strawson's argument about the irrationality of any attempt at universal abandonment of the reactive attitudes as a putative justification of praising and blaming (1990, pp. 19-20).
61. Honderich (1988, p. 449).
62. Ibid., pp. 526-7.

63. In fact Honderich is equivocal about justification. He suggests, for example, that those feelings of resentment and appreciation that are based on origination are unfair, because based on a falsehood (Honderich, 1988, p. 520). Since he also claims that the relationship between the feelings and belief in origination as their ground is not logical, but simply a matter of how we are disposed to think, he ought to have said that belief in determinism will tend to make us *feel* that such feelings are unfair.
64. Honderich (1988, p. 475). Cf. Nagel, who comments that 'the problem of free will is not in the first instance verbal. It is not a problem about what we are to say about action, responsibility, what someone could or could not have done, and so forth. It is rather a bafflement of our feelings and attitudes: a loss of confidence, conviction or equilibrium' (1986, p. 112).
65. Honderich (1988, pp. 406-9).
66. See Kane (1994, p. 49).
67. The Strawsonian explanation of our inability to persist in the reactive attitudes while adopting an objective attitude - such as thinking of someone as determined - also entails a denial of Honderich's claim that there is a family of reactive attitudes that we can persist in while thinking of someone as determined. Honderich suggests that if we consider the grounds of the feeling of vengefulness towards someone who has injured us (that they did not act out of character, or in ignorance, etc.), we can judge that they are consistent with determinism, and this will issue in the response of 'intransigence' about determinism (the response that determinism changes nothing). If instead we think that our feelings of vengefulness are directed at a corpus of dispositions including certain vicious desires, or if we consider that there is a causal explanation of those vicious desires, our feelings of vengefulness will be stymied (or at any rate some of them will). The two procedures for assessing the compatibility of the feelings with the thought of determinism are, however, quite different. In the first procedure we engage in an assessment of how certain grounds (those that can be expressed by saying that our adversary acted voluntarily) of our vengefulness stand in relation to determinism, and judge that there is no incompatibility. In the second, we think of the object of our vengefulness as a corpus of decisions, or his vicious desires as having a causal explanation, and suffer a loss of confidence in our vengefulness. The first procedure involves an intellectual examination of the grounds of feelings, whereas the second involves our testing out the feelings against certain thoughts associated with determinism. To support his claim that we have attitudes that are at home with determinism, in a sense that the other family of attitudes are not, Honderich would have to show that we can persist in *any* feelings of vengefulness, resentment, appreciation, etc. while thinking of their object as determined or as something other than a person. Honderich does claim that we would have strong feelings that are directed against 'a pair of vicious desires, within a certain corpus of dispositions' (1988, p. 407), as well as having feelings about their putative owner. But to say this is to suggest that the former feelings are discrete and discriminable from the latter. For my own part, I have no idea what it would mean or how it would feel to have feelings about someone's vicious desires which are distinct from any feelings I have about *him,* and therefore no idea how I could persist in the former feelings without the latter.
68. Honderich (1988, pp. 382-93).

69. Double (forthcoming: a). In reply to Double's arguments, Honderich (forthcoming) points out that although attitudes lack truth values, one of the two families of attitudes, in addition to various hopes, feelings and desires, does include images or propositional contents of an originator, which acts as the ground for those feelings and desires, and which is inconsistent with determinism.
70. Double offers three exemplars of what it is to be free: *the Reasonable Man, the Pure Rational Ego* and *the Nonegocentric Actor;* the first and second of these match the image of the free person as having a capacity for reflection and evaluation and an ability to act on it, with the third broadly corresponding to the kind of freedom exemplified in existential choices. Double's arguments for the irresolvability of the free will problem are further developed in his forthcoming *Metaphilosophy and Free Will* (forthcoming: b).
71. Allowing that it would make sense to talk of being unfair to mechanisms.
72. See, for example, Epicurus (1926) (c.300 BC), p. 113; Gtiinbaum (1971, pp. 309-10); Honderich (1988, pp. 360-75, 410-20); Hookway (1989). I criticise Honderich's argument that determinism gives cause for dismay about our actual or potential knowledge, in Magill (1992).
73. Kant famously takes the view that awareness of our ability to decide freely is dependent on our awareness of the moral law (1948, p. 4n). The claim is challenged by Galen Strawson (1986, pp. 65-6), and Double (1991, pp. 62-74).
74. Frankfurt (1978, p. 162); Watson (1987a, p. 168).
75. Reid (1846, pp. 608-13); Velleman (1992, p. 462).
76. Double (1991, p. 137).
77. I consider this argument in Chapter 3.
78. For similar arguments in support of the view that moral responsibility is not all that we want from free will, see Slote (1986, pp. 134ff).
79. Double has replied to this, in correspondence, that the argument only works if it can be shown that '(a) holding weak-willed persons responsible really *is* correct, and (b) weak-willed persons really *are* not free'. I do not think my argument does depend on this (although I think weak-willed persons are in a sense responsible and are in a sense unfree), since it is not intended as an endorsement of the Stoic view but only as evidence that it is possible to come to divergent judgements about moral responsibility and free will. It could of course be claimed that the Stoic view is so seriously wrong-headed that it cannot be taken as signifying anything about the relationship between free will and moral responsibility, but I doubt that Double could consistently argne for this, given his claim that the key terms and the attitudes that underlie them are incoherent.
80. What I am proposing is therefore different from the approach taken by Waller (1990), who distinguishes issues about free will from those about moral responsibility in order to advance a compatibilist account of the former along with a behaviourist denial of the latter.
81. Honderich (1988, p. 386).
82. G. Strawson (1986, pp. viii and 242).
83. Although Double does not think that our attitudes are incompatible with determinism, he does claim that there is a strong libertarian strain in our unreflective thinking about freedom.

84. A failing that Peter Strawson sought to remedy in 'Freedom and Resentment'.
85. Frankfurt (1971).
86. See Kane (1989, 1994).

Notes to Chapter 2: Moral Responsibility

1. Harry Frankfurt describes it as 'the principle of alternate possibilities' (1969), and Martha Klein as the 'C-condition' (1990).
2. Frankfurt (1969).
3. Ibid., p. 838.
4. P. F. Strawson (1982, pp. 62-7).
5. Klein (1990, pp. 8-9).
6. P. F. Strawson (1982, p. 79).
7. P. F. Strawson (1980, p. 265).
8. Cf. P. F. Strawson: 'Even when a formula [for justifying the reactive attitudes and practices] has been found ('contra-causal freedom' or something of the kind) there still seems to remain a gap between its applicability in particular cases and its supposed moral consequences' (1982, p. 79).
9. According to Nietzsche, for instance: 'The idea, now so obvious, apparently so natural, even unavoidable, that had to serve as the explanation of how the sense of justice ever appeared on earth - 'the criminal deserves punishment *because* he could have acted differently' - is in fact an extremely late and subtle form of human judgment and inference (Nietzsche, 1969, p. 63).
10. Spinoza (1677, I.iv).
11. J. Bennett (1980).
12. P. F. Strawson (1980, pp. 262-4).
13. Spinoza claimed, for example, that bad behaviour is often the product of *inadequate ideas* and *passive emotions.* A passive emotion is one that has us in thrall to illusion; involving confused beliefs. Cruelty, for instance, 'is the desire whereby any one is incited to work evil to one whom we love or whom we pity' (Spinoza, 1910, Part III, Def. xxxvm; see also Goodman, 1987, p. 133).
14. This point was put to me by David Cockburn. I suppose that incompatibilists might reply to this that in the absence of determinism, any agent would have the capacity at certain points to transcend those limitations and that failing to do so would mean that they are responsible for any bad action that results from such failure. The point, however, is that someone subject to the limitations of vision we are discussing would have no (subjective) reason for transcending them, and no cognitive means of doing so, regardless of whether her actions and choices are causally necessitated.
15. Perhaps there are Tory retributivists who have no need to ignore such feelings because they do not have them. In that case they lack something distinctively human, which places them beyond our discussion.
16. Bernard Williams, quoted by J. Bennett (1980, p. 25).
17. Ayer (1980, p. 12).
18. Non-cognitivists who reject the idea of foundational convictions are simply out of touch with ordinary moral thinking. It is ironic that they should regard Hume, who was very much in touch with ordinary moral thinking (without always being right about it), as any kind of inspiration for their views.

19. Those who question this should ask themselves whether they could ever bring themselves always to feel that it is right to forswear blame or resentment in response to bad or injurious behaviour. I develop a similar argument in relation to the justification of punishment in Magill (1996).
20. The idea that they are a bottom line appears to inform Watson's otherwise excellent discussion of the reactive attitudes in 'Responsibility and the Limits of Evil' (1987b).
21. Although the objective attitude can also be informed by sheer intellectual curiosity - an interest in knowing more - in which concern for the well-being of others plays no necessary part.
22. This simplistically assumes, of course, that what is implied by the objective attitude in therapy or medicine is a straightforward matter, and iguores large debates about therapist/client and doctor/patient relationships.
23. Hume (1975b, p. 229).
24. On the other hand, it is possible to have a principled indifference to the lot of some victims of crime: if they are rich, say, or if they are thought to benefit from the bad circumstances of others.
25. This is not to say that metaphysical attitudes about or related to determinism do not sometimes play a part in what the man or woman on the Clapham Omnibus thinks or says about punishment and blame. I do not think that non-philosophers, in being non-philosophers, are proof against philosophical confusion. My claim is that while metaphysical thoughts or attitudes may enter into the ways in which we perceive the difficulty about whether to blame or understand, the difficulty does not have its source in such attitudes.

Notes to Chapter 3: Free Will

1. See Kenny (1975), for a discussion of the 'liberty of spontaneity' and the 'liberty of indifference'.
2. Watson (1982a).
3. The seminal account of identification as consisting in having higher-order desires in respect of first- and lower-order desires is contained in Frankfurt's 'Freedom of the Will and the Concept of a Person' (1971). The case for evaluative judgements as the basis of identification is given by Gary Watson in 'Free Agency' (1982a). I discuss the differences between the two accounts in Chapter 6.
4. See also Don Locke's (1986) example of the airline pilot who is hijacked and instructed at gunpoint to fly to Cuba. As luck would have it, the pilot's lover is in Cuba and flying there is exactly what be wants to do. We would not describe his actions as any freer, however, than if he had had no desire at all to fly to Cuba.
5. *Because* is intended to indicate the agent's reasons for acting and not simply what causes him to act. Robin Taylor has suggested to me that it is possible that an agent might be mistaken about what be really wants, but be compelled by an external agency to do it anyway. If the external agency compels the agent to do what he really wants only because it is what be really wants, then this would, in a sense, entail that he does what be really wants because it is what he really wants. Although he would be doing what he really wants,

however, he would not be doing it for that reason and therefore he would not, in the relevant sense, be doing what he really wants *because* it is what he really wants. See also Chapter 6, n. 11.

6. It has been argued that free will cannot consist in doing what we really want, since it is possible that what we really want has been modified by neurosurgery, hypnosis or brainwashing (see, for example, Slote, 1986, p. 137; and Fischer, 1982, p. 37). As Daniel Dennett has pointed out, desires, values and beliefs can also be modified by reasoned argument without our being inclined to say that this in any way limits our free will. Providing we retain the capacity rationally to assess and evaluate our wants and values, why should it matter that some of them may have been introduced by peculiar means? If on the other hand we lack such a capacity, we will be unfree because of that and not because our wants and values have been induced in us. See Dennett (1981b, pp. 252-3); Double (1991, pp. 53-5); Watson (1987a, pp. 151-3). Cf. Mele (1995, Chapter 9).
7. See Double (1991, p. 38), for a useful summary of the personal characteristics that are required for self-determination.
8. When referring to the Stoic tradition, therefore, I will have in mind not only the Stoics themselves, but also those others who broadly share the Stoic conception of freedom. In addition to the Stoics, according to Adler, the tradition of 'the acquired freedom of self perfection' encompasses Plato, Plotinus, Augustine, Maimonides, Aquinas, Duns Scotus, Luther, Spinoza, Locke, Leibniz, Montesquieu, Rousseau, Kant, Hegel, Freud, Dewey, Whitehead and Russell, and arguably also Comte, Marx, Engels, Nietzsche and Durkheim (Adler, 1958, p. 134).
9. Richard Double (1991) has argued that since theories of free will of the kind we have been considering allocate a central position to rationality, they face difficulties in saying how rational our acts and decisions need to be to qualify as free. It conflicts with our understanding of free will to describe utterly irrational and capricious decisions as free, but if we set the normative requirement for rationality too high, our account of free will conflicts with prephilosophical notions of freedom and is therefore open to the charge of redefining it.
10. Epictetus's comments (which echo the view of wrongdoing as caused by limitations of vision that I discussed in Chapter I) are characteristic of the Stoic tradition: 'Who, then, wishes to live in error? No one. Who wishes to live deceived, impetuous, unjust, unrestrained, peevish, abject? No one. Therefore, there is no bad man who lives as he wills, and accordingly no bad man is free' (Epictetus, 1928 (*c.* AD 100), pp. 245-7).
11. Quoted in Adler (1958, p. 235).
12. Wolf (1987).
13. He could conceivably acquire the kind of 'external' sense of the distinction that a congenitally deaf person might develop in respect of musical harmony, through an understanding of mathematical relationships between note frequencies. What he would continue to lack, however, would be anything other than a prudential sense of why it is important to do what is right.
14. This argument was put to me by David Cockburn.
15. Wolf (1990).
16. Frankfurt (1975).

17. Ibid.
18. Ibid., p. 122.
19. Ibid., pp. 134-5.
20. See Slote (1986).
21. Whereas *free will* consists in doing what one *really* wants because it is what one *really* wants, the adverb *really* is inappropriate for freedom of action because the obstacles to freedom of action are external to the agent and therefore do not include conflicts of motive.
22. *External* and *internal* are perhaps not the happiest categories to use in drawing this distinction. Where, after all, should a brain tumour be placed in the division of freedom-restricting circumstances? I stick with *internal* and *external* here because of their familiarity, although more apt labels could be chosen, e.g., *circumstantial* and *motivational.*
23. Frankfurt (1975, p. 113).
24. Locke (1986, p. 99).
25. Ibid., p. 104.
26. See Frankfurt (1975); Locke (1986); Slote (1986).
27. Locke (1986, p. 102).
28. Adler (1958, p. 253). There is disagreement about this within the wider Stoic tradition, with philosophers such as Rousseau, Hegel and Dewey arguing that free will can only be acquired within an appropriate social framework (ibid., pp. 291-306). These thinkers also differ from other members of the tradition, therefore, in distinguishing self-determination from self-sufficiency and rejecting the latter as a chimera.
29. Slote (1986, p. 133).
30. See Chapter 2, n. 13.
31. Not forgetting of course that the Stoic advocacy of detachment is rather less saintly than Spinoza's.
32. Indeed this is in keeping with Spinoza's view that freedom consists in acting according to what is essential to one's nature, as opposed to what is inessential and contingent (Spinoza, 1910, Part I, Def. VII).
33. Wittgenstein (1958, 115).
34. See Dennett (1984), for many inventive assaults on this picture.
35. Double (1991); Wolf (1990, p. 74). See also Adler (1958, p. 573). It might be suggested that the rationality of a choice is a necessary but not sufficient condition for its being free, and that a further requirement is that an agent must have the capacity to choose otherwise. But if it is accepted that choosing otherwise than rationally is necessarily choosing unfreely, this suggestion is incoherent. If the ability to choose freely were to depend on an ability to choose unfreely, with what would we be contrasting the unfreedom of an unfree choice?
36. Double (1991, p. 108); Kane (1994); Klein (1990); G. Strawson (1986, p. vii).
37. Charles Taylor (1982). A revised version of this appeared as C. Taylor (1985a).
38. This example was suggested to me by Meena Dhanda.
39. Double (1991, p. 108).
40. See Norman (1983, chapter 3).
41. Quine suggests that 'free act' is a pleonasm, but offers no argument in support of this (Quine, 1987, p. 71).

42. It might be argued that one's capacity for choice could just as well be claimed as an attribute of self-determination - and thus of free will - as of agency, and therefore that there is no basis for a clear-cut distinction. Self-determination, however, would not exclude reasons doing the choosing, all by themselves, without some prior conception of what it is that is self-determining.
43. Kane (1994, p. 35). Cf. G. Strawson (1994).
44. Kane (1994, p. 37).
45. This is how Galen Strawson, for one, appears to understand the notion; although he also thinks that it should not be divorced from moral responsibility (1986, pp. vii-viii).
46. Double (1991, p. 108); Slote (1986, p. 138).
47. G. Strawson (1986, pp. 249-58).

Notes to Chapter 4: Can We Experience our Decisions as Caused?

1. Incompatibilists have argued that when we *act* we do so *under the idea of freedom* in the undetermined sense (notably Kant, 1948, pp. 107-23; but also Reid, 1846, p. 619, and van Inwagen, 1983, pp. 153-61): or at least that when we deliberate and decide we do so. I take this to be equivalent to the claim I am considering, that the *experience* of deliberating and deciding necessarily involves belief in undetermined freedom. If the idea enters into our experiences of deliberating, acting and deciding, then any argument about the veridicality of those experiences will have consequences for it. If the idea does not enter into those experiences, then there is no reason to suppose that we have to act or decide under it.
2. Van Inwagen claims that one who believes in determinism while continuing to deliberate and act is condemned to logical inconsistency (1983).
3. As it stands, this is incomplete. A complete analysis would refer to the experience of causes *producing* their effects. However, as I see it, this feature of the experience of causation does not reflect an objective characteristic of causation, and would not affect the conclusions I reach in this chapter. For reasons of economy, therefore, it has been excluded. My account of the way that productivity figures in the experience of causation is set out in 'Cause, Productivity and Experience' (unpublished paper).
4. Hume (1978, p. 170).
5. See Kneale (1974); Mackie (1975); von Wright (1974, pp. 7-10).
6. Hume (1978, p. 170). For a discussion of the relationship between Hume's two definitions of causation, see Lesher (1973).
7. Anscombe (1975); Ducasse (1974b); Humber and Madden (1974).
8. This account of the role of background conditions in the experience of causation fits Davidson's (1980a) account of the logic of singular causal statements as entailing the existence of a covering law without entailing any particular law, and that what justifies us in accepting singular causal statements is often evidence of regular connection between event types when conditions are favourable.
9. It would be tempting to assume that this difference might be accounted for as a projection onto a temporal distinction between background conditions being thought of as enduring and stable, as against the cause being a definite change and perhaps unusual or deviant (cf. Hart and Honore, 1974). While

this may be true of some contexts, however, in others what is experienced as a background condition could be experienced instead as the singular cause. Imagine you have just discovered that there has been a build-up of gas in a room, because of a leak, and that the central heating boiler, which is in the same room, is just about to switch on, having arrived at its preset time. If there is an explosion it would be natural to experience the later event - the switching on of the boiler - as the cause of it, even though the deviation from the normal, and what we would describe as the cause in an *explanation* of the explosion, would be the build-up of gas.

10. Ducasse (1974a, pp. 7-8). Elsewhere in the *Treatise*, Hume writes that the observation of resemblance between events *'produces* a new impression' in the mind, and of 'a *determination* of the mind to pass from one object to its usual attendant'; that this 'determination is the only *effect* of the resemblance; and therefore must be the same with power or efficacy, whose idea is *deriv'd* from the resemblance'; that 'Necessity ... is ... a detennination to *carry* our thoughts from one object to another'; and that it is 'that *propensity*, which custom produces, to pass from an object to the idea of its usual attendant' (1978, p. 165. My emphasis).
11. By contrast, to experience events or states as necessitated arguably does not require that they be experienced as caused. One who suffers from acute depression might be quite unable to believe that things will ever get better without necessarily experiencing this as being caused by any particular event or circumstances.
12. Two such philosophers are Anscombe (1975) and van Inwagen (1983, pp. 138-41). Cf. Honderich (1988, pp. 57-63). I deal with van Inwagen's arguments in 'What can God Show us?' (unpublished paper).
13. Gauthier (1967); Ginet (1970); Hampshire and Hart (1958); R. Taylor (1964).
14. This argument was put to me by Danny Goldstick. Cf. Ayer (1982); Hume (1978, pp. 410-12); Hobart (1966).
15. It might be argued that it would be her moral sense or her virtuousness that would cause her refusal, but these again are neither events nor singular.
16. Davidson (1980b).
17. A position with which, as I shall make clear below, I agree.
18. G. Strawson (1986, p. 70).
19. Ginet (1970); R. Taylor (1964).
20. For discussions of Ginet's arguments see Canfield (1962); Cowan (1969); Gauthier (1967); J. O'Connor (1967); Roxbee Cox (1963); Thalberg (1964).
21. Ginet (1970, p. 122).
22. Hampshire and Hart (1958, pp. 2-3).
23. Ibid., p. 2.
24. R. Taylor (1964, p. 75).
25. O'Shaughnessy (1980, pp. 300-2). See also Velleman (1989), who describes decisions as 'self-fulfilling predictions' (p. 91) and also 'self-fulfilling beliefs' (p. 95n).
26. Austin (1962).
27. Although this is not always so: see Velleman (1992, pp. 464-5), for an example of an unconscious intention.
28. O'Shaughnessy (1980, p. 301).
29. Ibid.

30. Sometimes a definite act of decision may be required even when a natural terminus has been reached. If deliberation unambiguously allows only one conclusion, but an undesirable one, one may resist acceptance of it for a time before resolving to be clear-headed about it.
31. Dennett (1984, p. 80).
32. Ginet (1970, p. 122).
33. R. Taylor (1964).
34. Pears (1975).
35. Ibid., p. 17.
36. The example is taken from Canfield (1962).
37. Robert Kane has suggested to me that most incompatibilists would argue that it is the experience of decisions *as free* that would be illusory if determinism were true. The addition of *as free* changes nothing, however, since it remains the case that if the claim that our experiences of decisions (free or otherwise) may be illusory is to mean anything, it can only be in contrast to logically possible veridical experiences. Since the supposed illusoriness of decisions, or of their being free, is thought to be entailed by their being caused or causally necessitated, the corresponding veridical experiences would in either case be ones that are experienced as caused or causally necessitated.
38. A second-order desire in Frankfurt's sense.
39. Popper (1950); Ryle (1949); MacKay (1960); Oldenquist (1964).
40. Oldenquist (1964).
41. This is not to say that an agent could not accurately identify elements of her deliberative processes: only that she could not identify them as causes; or at least not where there was any difficulty in deliberating.
42. Of course the incompatibilist argument was that if determinism is true our experiences of deliberating and deciding are illusory, not because they would be caused, but because they would be *causally necessitated.* But if difficult decisions logically cannot be experienced as caused, then neither can they be experienced as causally necessitated.
43. Honderich (1988, pp. 382-93).
44. MacKay (1960).
45. Mark Bernstein has suggested to me that knowing that such a state of affairs exists entails only that nothing you do will be inconsistent with its existing. But if you do not know that nothing you do will be inconsistent with the existence of the state of affairs, while you might have a true belief that such a state of affairs exists, you could not be said to have knowledge of it. Therefore, if knowledge that such a state of affairs exists were possible, it would entail knowledge that nothing one could do would be inconsistent with its existing.

Notes to Chapter 5: What are Actions?

1. Reid (1846, pp. 609-11).
2. Ibid.
3. Nagel (1986, pp. 110-11).
4. Such a view, or something like it, seems to inform Richard Taylor's recantation of agent-causationism: see R. Taylor (1982, especially p. 227).

5. Assuming that raising my arm is all I am doing, and not, for example, voting.
6. The kind of weak-willed behaviour involved in lacking the resolution to get out of bed will be touched on in both chapters.
7. See Austin (1961b); Cavell (1976); Ryle (1949).
8. Austin (1961b).
9. Hobbes (1962, p. 47). The contrast Hobbes draws between 'animal motions' and 'vital motions' (breathing, circulation, digestion, etc.) will not do as it stands, however: for example, under which category should epileptic fits be included?
10. *Physics,* 256a, in Aristotle (1984) (c.350 BC). Aristotle's intention is to distinguish man as 'a mover that is not so in virtue of being moved by something else'.
11. This does not require that a particular individual has to be explicitly identified in any description of an action, as we can see from descriptions like 'the bomb had been set to go off'.
12. If it is intelligible to think of real and imagined non-organic entities being conscious (e.g., Martians, robots and computers), then it might also make sense to extend the distinction to their behaviour.
13. Davidson (1980c).
14. Dennett (1976 and 1981b).
15. Dennett (1981b, p. 238).
16. Cf. Dennett (1976). Dennett's argument that animal actions are just one form of intentional behaviour, differing in detail and complexity, but not, *qua* intentional, different in kind from the behaviour of trees and thermostats, depends on the idea that ascriptions of intentional states are made to the individual and not to the precise mechanisms that physically cause the behaviour so explained. In contrast to Dennett's view, the account of actions I will advance involves intentional states acting directly on the mechanisms that cause actions and therefore playing an ineliminable role in any explanations of the initiation of actions.
17. It is worth remembering that philosophers who believe that engaging in laboured definition of key terms is the be-all-and-end-all of good philosophy have no less a tendency to fall back on questionable intuitions when definitional analysis produces no substantive conclusions.
18. Hobbes (1962, pp. 47-58); Davidson (1980c); Goldman (1976).
19. Davidson (1980c, p. 77).
20. J. D. Velleman (1992, p. 462).
21. Ibid., pp. 462-3.
22. D. Bennett (1965).
23. For a discussion of the problem of 'external causal deviancy' and suggestions about how it can be overcome, see Bishop (1989, pp. 125-32).
24. Davidson (1980c, p. 79).
25. Frankfurt (1978).
26. Ibid., p. 158.
27. Ibid., p. 160.
28. Ibid.
29. Bishop (1989, p. 171); Churchland (1986, pp. 430-1).
30. Frankfurt (1976, p. 161).

31. Ibid.
32. Chisholm (1976, 1978 and 1982); R. Taylor (1966).
33. Chisholm (1982, p. 32).
34. Chisholm (1978, p. 623). The view of agency we all unreflectively share also extends to the actions of animals, but it would be odd to think of them as unmoved movers.
35. Chisholm (1982, p. 32). Chisholm has subsequently abandoned agent-causationism (see R. J. Bogdan, 1986, pp. 214-15).
36. R. Taylor (1966, pp. 103-6, 122-33, 140).
37. Ibid., pp. 114-15; Watson (1987a, pp. 164-9).
38. See Chapter 1, n. 37.
39. I do not mean to imply by this that we are always directly aware of being somehow involved in our actions. There are many things we do that involve subsidiary actions which we carry out without thinking of or being especially conscious of (e.g., peddling and steering while cycling). My point is that to the extent that we are aware of or focused on an action we are performing, we experience it as intimately involving us; that we never experience our actions as discrete from us.
40. One feature of which is its inability to account for the timing of actions. We think of the timing of an action as being under an agent's control, but since agent-causation is not understood to be identical to an event, it denies us any possibility of saying how or why an agent acts when she does; or indeed of understanding how agent-causation can exist within the temporal order: a problem which Kant, in contrast to Chisholm and Taylor, was well aware of.
41. It might be thought that, if the relationship between agent and action is not experienced as causal, this rules out the possibility, which I allowed in the previous chapter, that some decisions may be experienced as caused. My argument, however, is not that we cannot experience our actions as caused, but rather that we do not experience our relationship to them as causal.
42. Davidson (1980c).
43. Ginet (1990).
44. See Goldman (1976, pp. 81-2), for interesting suggestions about what causes the actish quality, or as Goldman describes it 'the feeling of voluntariness'.
45. Often I will have the desire to be up and about, but then be distracted by some memory or thought. The temptation here is to say that it is the distraction that causes me not to act on my desire, but isn't it just as likely that the distraction is a consequence of my failure to get up rather than a cause of it?
46. It might be thought that smoking in order to fill out the time is caused by a desire to be doing something (with one's hands, perhaps). But smokers are sometimes able to reflect on such desires and to judge that they are weak or that they don't have them: that the supposed desires are deceptions borne of rationalising the inclination to smoke. Cf. Mele (1992, pp. 79-85), on the plasticity of motivational strength.
47. In small children and those others Frankfurt calls wantons, we can say that the Will is triggered by whatever occurrent desires, fears, etc. are strongest. A person with free will, by contrast, would have a will that has been habituated via training, practice, calculation, reflection, and so on to be responsive

to her higher-order preferences and values. Those, other than wantons, who lack free will could be persons who have repeatedly behaved in ways that have habituated their wills to be triggered by desires or values that are not what they really want. What it means to say that someone does not really want to do what she does remains to be settled: it is the problem of *identification,* and I shall return to it in the following chapter.

48. I discuss this in the following chapter.
49. We might add the qualification here that 'A believes x-ing is a way to bring about ϕ_i' should be understood as stating that *A's* believing this involves *A* having an understanding of how x-ing will bring about ϕ_i. The point being to exclude the kinds of 'external' deviant causal chains considered earlier.
50. This clause is suggested by Davidson (1980c), who points out that a person may want ϕ_2 and believe x-ing is the best way to bring about ϕ_2, but still fail to reason out that this implies that *x* is worth doing, and therefore fail to *x*.
51. The inclusion of 'appropriately' in the analysis is intended to exclude cases where the Will has been conditioned in such a way as to cause behaviour that does not realise the agent's wants and beliefs. Cases of this sort would fall within the class of unintentional actions and are dealt with below. It might be thought that an analysis of a causal relationship should not include any reference to causes, and therefore that the implicit reference to appropriate causation by the Will vitiates the analysis. However, the analysis is of the causal relationship between attitudes and actions, of which one element is the operation of a mechanism. Provided that in principle a causal analysis of the operation of the mechanism could be given, there is nothing more objectionable in this than, for example, 'If I turn the key, and the engine is operating normally, then the car starts'.
52. See Ginet (1990, pp. 15-22).
53. Ibid., p. 3.
54. It is possible to imagine circumstances in which one might have an illusory experience of this: because the causal operation of one's will is blocked by *Q* signals from Mars, but Z signals from Venus cause one's body to behave just as if one's will were operating normally (Mele, 1992, pp. 248-9). The actish quality is not absolute proof against illusion, therefore, but under normal circumstances it is no more suspect than our experiences of the world around us.
55. See Bishop (1989, pp. 137-40), for a useful discussion of such approaches.
56. Morton (1975); Peacocke (1979, pp. 66-71, 79-86).
57. Nevertheless, it is always possible to find a philosopher who will attempt to deny the undeniable. Michael Zimmerman argues that there are 'clear cases of sentences in which a proposition involving no action or actions features as the object of someone's intention. Witness "Smith intends that Jones should be happy," "I intend that Smith's car should be waiting in front of the bank at 3 p.m. sharp," and so on' (Zimmerman, 1984, p. 200). Clearly, however, although such sentences contain no explicit references to actions, they do make elliptical reference to actions the speaker will perform, and imply that such actions are intended to bring about the desired states of affairs. If the speaker does not intend to do anything to bring about the desiderata, she can only hope that Smith's car will be there on time or that Jones will be happy.

58. By contrast, since the Will is taken to be a set of behaviour-causing mechanisms that are exclusively triggered by intentional states its definition does not presuppose the concept of action.
59. Bishop (1989, pp. 95-8).
60. Ibid., pp. 189-90.
61. See Sellars (1976, p. 47); and also Goldman (1976).
62. Incidentally, the absence of prior intention in the two examples does not imply that the actions described were unintentional. The intentions with which the agents acted would have been given by whatever intentional states (love, strong desire, or whatever, together with beliefs about how to satisfy them and that it was possible to do so) it was that caused them to act.
63. Responsiveness to context/situation is required for sustained and non-basic actions (i.e. actions whose description includes the intended outcome of the agent's basic bodily action), and is a requirement of their being actions, but is not required by a satisfactory analysis of actions as such. Responsiveness to context, therefore, addresses an issue that is logically secondary to the problem of action with which we are concerned.
64. Ginet (1990, p. 11).
65. Ibid., p. 10.
66. The two cannot be identical, since character is also taken to refer to a person's characteristic moods and behaviour, as well as being the source of a person's actions.
67. Davidson (1980d).
68. See, for example, Davidson (1980e) and Bishop (1989, pp. 110-20). See Watson (1977), for a dissenting view.
69. Davidson (1980, p. xii); Bishop (1989, p. 111).
70. For those who do not know it, the game is played between two players, each of whom holds one hand behind his or her back and who then simultaneously bring the hand forward in a shape representing, in the simplest version of the game, paper, stone or scissors. Stone trumps scissors (because scissors will be broken or blunted by a stone), scissors trump paper (obviously) and paper trumps stone (because, if memory serves, the stone is imagined to be blunt and paper can be wrapped round it - it is only a game!). Thus, the effectiveness of any shape depends not only on the nature of the object it represents but also on that of the object represented by the shape it acts against.
71. There is an interesting parallel here between this account of motivation as a compromise or conjuncture between rational reflection and the influence of past behaviour and conservative ideas about social development as a compromise between rational thought and unplanned social evolution. The compatibility of these ideas clearly shows up in Hume.
72. Bishop (1989, p. 119).
73. This would be something like the kind of 'all-out judgment' described by Davidson (1980d, p. 99), although unlike him I do not think that such judgements are identical with either actions or intentions to act, and neither do I think that they are about what is *best* (most rational) to do. An army officer's subordinates will understand his orders as indicating what is to be done, rather than what it is best to do, even though they may have good reason to believe that the orders are for the best.

74. My account might not satisfy the details of any and every incompatibilist argument about dual-rationality and dual-control, but as I have argued before, what is important about incompatibilist arguments, and what gives them their plausibility, is the everyday unreflective understandings, senses and experiences of acting and deciding they have sought to give expression to, and which compatibilists have repeatedly failed to do justice to. If we can give better expression to those understandings and experiences, without recourse to incompatibilist theories and concepts, there is no further need, for purposes of argument and persuasion, to answer every detail of their claims and arguments. For a summary of recent arguments about dual-control and rationality, see Double (1991, Chapter 8).
75. If, instead, the manipulator were to trigger the agent's will in just the way it would have been triggered without his modifications, and there is no reason to suppose that he would do otherwise, his intervention would make no difference to how she experiences her actions, but neither would it matter to their actually being free actions. All his intervention would achieve would be to create an unnecessary additional step between her motives and her will. Provided the extra step could be relied upon to make no difference to how her will responds to her motives and intentions, it should not be regarded as affecting her ability to act in any way.
76. Richard Taylor objects to any analysis of the role of the agent in terms of events or states, which might be induced by 'electrical impulse, drug, or whatnot' (1966, pp. 94-5), that since such an event can be brought about without one's participation, what it causes cannot be 'anything I performed or was in any way responsible for' (ibid., p. 94). He also suggests that if an agent's causal contribution to her action is analysed in terms of an event or state, that state or event must also be considered her doing 'if the analysis is to have any plausibility at all' (ibid., p. 95), which could not be the case if it were caused by a drug or whatever. But there is no reason to accept that if any agent's contribution to her action is analysed in terms of causation by an event, that event must also be considered to be her doing. Taylor's argument flatly contradicts his own acceptance that agents can sometimes be causally necessitated to act as they do.

Notes to Chapter 6: Free Agency

1. See Davidson (1980e, pp. 35-6).
2. Reid (1846, p. 611).
3. Velleman (1992).
4. I have already argued that causation by intentions is not an essential feature of actions, but since the weighing up of reasons or motives that is involved in the kind of full-blooded agency I am attempting to specify normally culminates in a decision - even if immediately followed by an action - and since decision, as I argued in Chapter 4, necessarily involves the formation of an intention, then causation by intentions will be a normal feature of full-blooded agency.
5. Velleman (1992, p. 462).
6. Reid (1846, pp. 608-13).
7. Velleman (1992, p. 462).

8. Frankfurt (1971, p. 16), quoted in Velleman (1992, p. 463).
9. Velleman (1992, p. 471).
10. Frankfurt (1971). See also Double (1989); Lehrer (1980); Velleman (1992); Watson (1987a).
11. The concept of *identification* is, as I suggested in Chapter 3, central to the idea of being able to do what one really wants because it is what one really wants. It has been suggested that if free will consists in being able to do what one really wants because it is what one really wants, an agent who is mistaken about what she really wants might be compelled by some external agency to act in accordance with what she really wants and therefore that she would be 'compelled to act freely' even though she would not recognise what she does as acting freely (see Chapter 3, n. 5). What it means to say that an agent is mistaken about what she really wants is, presumably, that she has a desire or value of which she is unconscious or which she refuses to acknowledge, or that her acknowledged desires and values have implications of which she is unaware. But for as long as an agent is unaware of or refuses to acknowledge any unconscious desire or value, or any implication of her desires and values, her attitude towards them will be ambiguous and she will be unable to identify with them. If she is unable to identify with them, she cannot be free in acting on them.
12. Frankfurt (1971, p. 16).
13. Watson (1987a, p. 149); see also Watson (1982a).
14. Frankfurt (1971, pp. 16-17; and 1976, pp. 248-51).
15. Velleman (1992, pp. 472-4). Frankfurt concedes that he finds the nature of decision 'very obscure' (1976, p. 251).
16. Watson (1982a, p. 110).
17. Watson (1987a, p. 150).
18. Velleman (1992, p. 472n).
19. Ibid., p. 477.
20. Including the second-order desire usually to be motivated by the desire to gain the qualification, and, following her reflections, a desire that on this occasion her first-order desire to take it easy should motivate her actions.
21. Bishop (1989, p. 109).
22. Velleman (1992); Watson (1987a).
23. Velleman (1992, p. 472n).
24. C. Taylor (1985a).
25. Velleman (1992, p. 472).
26. Watson also rejects his previous view on the grounds that 'it conflates valuing with judging good. Notoriously, judging good has no invariable connection with motivation, and one can fail to "identify" with one's evaluational judgments' (1987a, p. 150). If one feels alienated from an evaluational judgement, however, it can only be because it is at odds with what one values more (in some respect), and what one values more may not be owed to one's sense of morality.
27. Watson (1987a, p. 150).
28. Cf. Watson (1982a, p. 104), on desires that temporarily influence an agent's evaluations.
29. Watson (1987a, p. 150).
30. Ibid.

31. This point is made by Charles Taylor (1982, p. 119) in criticism of Sartre and non-cognitivist moral philosophers who argue that agents choose their most fundamental values.
32. Although it may be immediately brought about by an action, for example, in those cases I mentioned in Chapter 4, where one makes a judgement that brings one's deliberations to a conclusion because one lacks the time to pursue them further.

    It is true that agents sometimes do reject values that hitherto have had a powerful influence on their actions and choices, but their doing so is informed by a prior realisation that they no longer wish to be motivated by the rejected values, and this in turn involves a realisation that those values are in conflict with more deeply held values. Even where there is a decisive act of rejection, therefore, in a sense the act of rejection is simply a confirmation or affirmation of what is already the case.
33. Velleman (1992, p. 479).
34. Ibid., p. 477.
35. Ibid., p. 479.
36. As I argued in the previous chapter, the relative strengths of reasons for acting is partly determined by practical judgements, strengths of desire and what desires the Will is most responsive to.
37. Velleman (1992, p. 478).
38. Velleman has replied to this, in correspondence, that the agent's functional role is not specified by whatever role the desire to act in accordance with reasons happens to play, but that it is independently specifiable and that the desire to act in accordance with reasons is what ordinarily plays this role. We may say that the functional role of the agent is that of being able to critically assess opposing motives and courses of action, and of being able to bring her behaviour in line with whatever judgement such assessment results in. Setting aside my argument that it is not a desire to act in accordance with reasons that motivates the process of practical thought, if a critical assessment of motives and action is prevented or perverted by external interference, without affecting the causal operation of the desire, doesn't this indicate that there must be more to the agent's functional role than what can be satisfied by the causal operation of that desire? It might be argued that if the process of critical assessment of motives is perverted, then the desire to act in accordance with reasons has not been satisfied, but if the desire to act in accordance with reasons fulfils the agent's functional role in virtue of its causal influence, that is a separate matter from whether its propositional content is satisfied.
39. Not all inferences are carried out actively. If they were, then of course the analysis would succumb to infinite regress.
40. For example, suppose someone is figuring out a plan of action, but while thinking it through she begins to feel repulsed by the thought that what is motivating her reflections is a desire to gain some small revenge on an acquaintance, and she abandons her scheming. Giving up her scheming is therefore motivated by the desire to do what she cares most about.
41. See O'Shaughnessy (1980, p. 301), and also my discussion of this in Chapter 4.
42. It would be rather as if one's actions were subject to the control of a

manipulator who sometimes allowed one to do what one wanted and sometimes did not.

43. Nevertheless, for some actions, whatever it would take to condition the Will to make them possible for an individual may be beyond her imagination, skill, intellect, character, or other resources (money, friends, family, education, social provision, etc.). Individuals like this will not only be unable to perform such actions, but will be *utterly unable* to do so. Since their inabilities are conditional on their share of the common wealth, we can fairly say that the capacity to act freely, as well as the opportunity to do so, is unequally distributed. Some people have more in the way of free agency than others: it is both conceivable and desirable that social arrangements be so ordered that those with less free agency are enabled to have much more of it.

44. If our actions were only made probable rather than causally necessitated by our antecedent mental states, it might be argued that agency is only partly illusory, since we can have an intelligible degree of control over outcomes that have a high degree of probability.

References

Adler, M. (1958) *The Idea of Freedom* (New York: Doubleday).

Anscombe, G. E. M. (1975) 'Causality and Determination', in Sosa (1975).

--(1976) 'Soft Determinism'. in G. Ryle (ed.), *Contemporary Aspects of Philosophy* (Stocksfield: Oriel).

Aristotle (1984) (c.350 BC), *The Complete Works of Aristotle,* vol. 1, ed. J. Barnes (Guildford: Princeton University Press).

Aune, B. (1982) 'Hypotheticals and "Can": Another Look', in Watson (1982).

Austin, J. L. (1961) *Philosophical Papers,* ed. J. 0. Urmson and G. J. Warnock (Oxford: Clarendon).

--(1961a) 'Ifs and Cans', in Austin (1961).

-- (1961b) 'A Plea for Excuses'. in Austin (1961).

--(1962) *How to Do Things with Words,* ed. J. 0. Urmson (London: Harvard University Press).

Ayer, A. J. (1963) 'Fatalism'. in *The Concept of a Person and Other Essays* (London: Macmillan).

-- (1980) 'Free-will and Rationality'. in van Straaten (1980).

--(1982) 'Freedom and Necessity', in Watson (1982).

Beauchamp, T. L. (ed.) (1974) *Philosophical Problems of Causation* (Encino: Dickenson).

Bennett, D. (1965) 'Action, Reason and Purpose', *Journal of Philosophy,* 62, 85-95.

Bennett, J. (1971) *Locke, Berkeley, Hume: Central Themes* (London: Oxford University Press).

--(1980) 'Accountability', in van Straaten (1980).

Berofsky, B. (ed.) (1966) *Free Will and Detenninism* (New York: Harper).

Bishop, J. (1989) *Natural Agency* (Cambridge: Cambridge University Press).

Bogdan, R. J. (ed.) (1986) *Roderick M. Chisholm* (Dordrecht: Reidel).

Brand, M. N. and Walton, D. (eds) (1976) *Action Theory* (Dordrecht: Reidel).

Canfield, J. (1962) 'Knowing about Future Decisions', *Analysis,* 22, 127-9.

Cavell, S. (1976) *Must We Mean What We Say?* (Cambridge: Cambridge University Press).

Chisholm, R. W. (1964) 'J. L. Austin's Philosophical Papers', *Mind,* LXXIII, 289, 1-26.

-- (1976) *Person and Object: A Metaphysical Study* (London: Allen and Unwin).

-- (1978) 'Comments and Replies', *Philosophia,* 597-636.

-- (1982) 'Human Freedom and the Self', in Watson (1982).

Churchland, P. S. (1986) *Neurophilosophy: Toward a Unified Science of the Mindi Brain* (London: MIT Press).

Collingwood, R. G. (1974) 'Three Senses of the Word "Cause"', in Beauchamp (1974).

Cowan, J. L. (1969) 'Deliberation and Determinism', *American Philosophical Quarterly,* 6, 1.

Davidson, D. (1980) *Essays on Actions and Events* (Oxford: Clarendon).

-- (1980a) 'Causal Relations', in Davidson (1980).

--(1980b) 'Actions, Reasons and Causes', in Davidson (1980).
--(1980c) 'Freedom to Act', in Davidson (1980). First appeared in Honderich (1973).
--(1980d) 'Intending', in Davidson (1980).
-- (1980e) 'How is Weakness of the Will Possible?', in Davidson (1980).
Dennett, D. C. (1976) 'Conditions of Personhood', in Rorty (1976). Reprinted in Dennett (1981).
-- (1981) ***Brainstonns*** (Brighton: Harvester).
-- (1981a) 'On Giving Libertarians What They Say They Want', in Dennett (1981).
-- (1981b) 'Mechanism and Responsibility', in Dennett (1981). Also included in Watson (1982).
-- (1984) ***Elbow Room: The Varieties of Free Will Worth Wanting*** (Oxford: Clarendon).
Double, R. (1989) 'Puppeteers, Hypnotists and Neurosurgeons', ***Philosophical Studies,*** 56, 163-73.
-- (1991) ***The Non-Reality of Free Will*** (Oxford: Oxford University Press).
--(1993) 'The Principle of Rational Explanation Defended', ***Southern Journal of Philosophy,*** XXXI, 2, 133-42.
--(1994) 'How to Frame the Free Will Problem', ***Philosophical Studies,*** 75, 149-72.
-- (forthcoming: a) 'Honderich on the Consequences of Determinism'.
-- (forthcoming: b) ***Metaphilosophy and Free Will*** (New York: Oxford University Press).
Ducasse, C. J. (1974a) 'Critique of Hume's Conception of Causality', in Beauchamp (1974).
-- (1974b) 'Analysis of the Causal Relation', in Beauchamp (1974).
Dworkin, R. (ed.) (1970) ***Detenninism, Free Will and Moral Responsibility*** (Englewood Cliffs: Prentice-Hall).
Epictetus (1928) (c. AD 100), ***The Discourses,*** vol. II, trans. W. A. Oldfather (London: Heinemann).
Epicurus (1926) (c.300 BC), ***The Extant Remains,*** ed. C. Bailey (Oxford: Clarendon).
Fischer, J. M. (1982) 'Responsibility and Control', ***Journal of Philosophy.***
-- (ed.) (1986) ***Moral Responsibility*** (Ithaca: Cornell University Press).
-- (1994) ***The Metaphysics of Free Will*** (Oxford: Blackwell).
Frankfurt, H. (1969) 'Alternate Possibilities and Moral Responsibility', ***Journal of Philosophy,*** 829-39. Also in Frankfurt (1988).
--(1971) 'Freedom of the Will and the Concept of a Person', ***Journal of Philosophy,*** LXVIII, 1, 5-20. Also in Frankfurt (1988).
-- (1975) 'Three Concepts of Free Action', ***Proceedings of the Aristotelian Society,*** 113-25. Also in Frankfurt (1988).
-- (1976) 'Identification and Externality', in Rorty (1976). Also in Frankfurt (1988).
--(1978) 'The Problem of Action', ***American Philosophical Quarterly,*** 15, 2, 157-62. Also in Frankfurt (1988).
--(1988) ***The Importance of What We Care About*** (Cambridge: Cambridge University Press).
Gauthier, D. P. (1967) 'How Decisions are Caused', ***Journal of Philosophy,*** LXIV, 5, 147-51.

Ginet, C. (1966) 'Might we Have No Choice?', in Lehrer (1966).
-- (1970) 'Can the Will be Caused?', in Dworkin (1970).
--(1990) *On Action* (Cambridge: Cambridge University Press).
Goldman, A. I. (1976) 'The Volitional Theory Revisited', in Brand and Walton (1976).
Goldstick, D. (1979) 'Why we Might Still Have a Choice', *Australasian Journal of Philosophy, 51,* 4.
Goodman, L. E. (1987) 'Determinism and Freedom in Spinoza, Maimonides, and Aristotle: A Retrospective Study', in Schoeman (1987).
Grilnbaum, A. (1971) 'Free Will and the Laws of Human Behaviour', *American Philosophical Quarterly,* 8, 4.
Hampshire, S. and Hart, H. L. A. (1958) 'Decision, Intention and Certainty', *Mind,* LXVII, 265, 1-12.
Hart, H. L. A. and Honore, A. M. (1974) 'The Analysis of Causal Concepts', in Beauchamp (1974).
Hobart, R. E. (1966) 'Free Will as Involving Determination and Inconceivable without It', in Berofsky (1966).
Hobbes, T. (1962) (1651), *Leviathan* (New York: Collier Macmillan).
Honderich, T. (ed.) (1973) *Essays on Freedom of Action* (London: Routledge and Kegan Paul).
-- (1988) *A Theory of Determinism: The Mind, Neuroscience, and Life Hopes* (Oxford: Clarendon).
-- (forthcoming) 'Compatibilism, Incompatibilism, and the Smart Aleck'.
Hookway, C. (1989) 'The Epicurean Argument: Determinism and Scepticism', *Inquiry,* 32, 1.
Humber, J. and Madden, E. H. (1974) 'Nonlogical Necessity and C. J. Ducasse', in Beauchamp (1974).
Hume, D. (1975) (1777), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals,* ed. L. A. Selby-Bigge (Oxford: Clarendon).
-- (1975a) *An Enquiry Concerning Human Understanding,* in Hume (1975).
-- (1975b) *An Enquiry Concerning the Principles of Morals,* in Hume (1975).
-- (1978) (1739), *A Treatise of Human Nature,* ed. L. A. Selby-Bigge (Oxford: Clarendon).
Kane, R. (1989) 'Two Kinds of Incompatibilism', *Philosophy and Phenomenological Research,* L, 2 (1989).
-- (1994) 'Free Will: The Elusive Ideal', *Philosophical Studies, 15,* 25-60.
Kant, I. (1948) (1785), 'Groundwork of the Metaphysic of Morals', in H. J. Paton (ed.), *The Moral Law* (London: Hutchinson).
--(1956) (1788), *Critique of Practical Reason,* trans. L. W. Beck (London: Collier Macmillan).
Kapitan, T. (1994) 'Critical Study of Richard Double's *The Non-Reality of Free Will', Noas,* 28, I, 90-5.
Kenny, A. (1975) *Will, Freedom and Power* (Oxford: Blackwell).
Klein, M. (1990) *Determinism, Blameworthiness and Deprivation* (Oxford: Clarendon).
Kneale, W. (1974) 'Natural Laws and Contrary-to-fact Conditionals', in Beauchamp (1974).
Lamb, J. W. (1977) 'On a Proof of Incompatibilism', *Philosophical Review,* LXXXVI, 1.

Lehrer, K. (ed.) (1966) *Freedom and Determinism* (New York: Random House).
-- (1980) 'Preferences, Conditionals and Freedom', in van Inwagen (1980).
-- (1982) 'Cans without Ifs', in Watson (1982).
Lesher, J. H. (1973) 'Hume's Analysis of "Cause" and the "Two-Definitions" Dispute', in *History of Philosophy,* II, 387-92.
Lewis, D. (1981) 'Are We Free to Break the Laws?', *Theoria,* XLVII, 3, 113-21.
Locke, D. (1986) 'Three Concepts of Free Action: I', in Fischer (1986).
MacKay, D. M. (1960) 'On the Logical Indeterminacy of a Free Choice', *Mind,* 69.
Mackie, J. (1975) 'Causes and Conditions', in Sosa (1975).
Magill, K. (1992) 'Epicurus, Determinism and the Security of Knowledge', *Theoria,* LVIII, 183-96.
-- (1996) 'The Idea of a Justification for Punishment', in *Contemporary Political Studies,* vol. 3: *Proceedings of the Annual Conference of the Political Studies Association,* ed. I. Hampsher-Monk and J. Stanyer (Oxford: Blackwell).
Mele, A. (1992) *Springs of Action* (Oxford: Oxford University Press).
-- (1995) *Autonomous Agents: From Self-Control to Autonomy* (Oxford: Oxford University Press).
Moore, G. E. (1912) *Ethics* (London: Thornton Butterworth).
Morton, A. (1975) 'Because He Thought He Had Insulted Him', *Journal of Philosophy,* 72, 5-15.
Nagel, T. (1986) *The View from Nowhere* (Oxford: Oxford University Press).
Nathan, N. M. L. (1992) *Will and World: A Study in Metaphysics* (Oxford: Clarendon).
Nietzsche, F. (1969) *On the Genealogy of Morals,* collected with *Ecce Homo,* ed. W. Kaufmann (New York: Vintage).
Norman, R. (1983) *The Moral Philosophers* (Oxford: Clarendon).
Nowell-Smith, P. H. (1960) 'Ifs and Cans', *Theoria,* XXVI, 2, 85-101.
Nozick, R. (1981) *Philosophical Explanations* (Oxford: Clarendon).
O'Connor, J. (1967) 'How Decision are Predicted', *Journal of Philosophy,* LXIV, 5, 429-30.
O'Connor, T. (1993) 'Indeterminism and Free Agency: Three Recent Views', *Philosophy and Phenomenological Research,* LIII, 3.
Oldenquist, A. (1964) 'Causes, Predictions and Decisions', *Analysis.*
O'Shaughnessy, B. (1980) *The Will: A Dual-Aspect Theory* (Cambridge: Cambridge University Press).
Peacocke, C. (1979) *Holistic Explanation: Action, Space, Interpretation* (Oxford: Clarendon).
Pears, D. (1975) 'Predicting and Deciding', in *Questions in the Philosophy of Mind* (London: Duckworth).
Popper, K. R. (1950) 'Indeterminism in Quantum Physics and Classical Physics, Part I', *British Journal for the Philosophy of Science,* I, 2, 117-95.
Quine, W. V. O. (1987) *Quiddities* (Cambridge Mass.: Harvard University Press).
Reid, T. (1846) (1788), 'Essays on the Active Powers of the Human Mind', in *The Works of Thomas Reid* (London: Longmans).
Rorty, A. E. (ed.) (1976) *The Identities of Persons* (Berkeley: University of California Press).
Roxbee Cox, J. W. (1963) 'Can I Know Beforehand What I am Going to Decide?', *Philosophical Review,* LXXII, 88-92.
Ryle, G. (1949) *The Concept of Mind* (Harmondsworth: Penguin).

Schoeman, F. (ed.) (1987) *Responsibility, Character and the Emotions* (Cambridge: Cambridge University Press).
Sellars, W. (1976) 'Volitions Reaffirmed', in Brand and Walton (1976).
Slote, M. (1982) 'Selective Necessity and the Free Will Problem', *Journal of Philosophy,* LXXIX, 1, 5-24.
-- (1986) 'Understanding Free Will', in Fischer (1986).
Sosa, E. (ed.) (1975) *Causation and Conditionals* (Oxford: Oxford University Press).
Spinoza, B. (1677) *Tractatus Politicus,* collected with *Tractatus Theologico-Politicus,* trans. R. H. M. Lewes (London: Routledge).
-- (1910) (1677), *Ethics* (London: Dent).
Strawson, G. (1986) *Freedom and Belief* (Oxford: Clarendon).
-- (1994) 'The Impossibility of Moral Responsibility', *Philosophical Studies.* 75, 1-2, 5-23.
Strawson, P. F. (1980) 'Reply to Ayer and Bennett', in van Straaten (1980).
--(1982) 'Freedom and Resentment', in Watson (1982). First appeared in the *Proceedings of the British Academy,* XLVIII (1962), 1-25.
Taylor, C. (1982) 'Responsibility for Self, in Watson (1982).
-- (1985) *Human Agency and Language* (Cambridge: Cambridge University Press).
-- (1985a) 'What is Human Agency?', in C. Taylor (1985).
Taylor, R. (1963) 'A Note on Fatalism', *Philosophical Review,* LXXII, 497-9.
--(1964) 'Deliberation and Foreknowledge', *American Philosophical Quarterly,* 1, 1, 73-80.
-- (1966) *Action and Purpose* (Englewood Cliffs: Prentice-Hall).
-- (1982) 'Agent & Patient: Is There a Distinction?', *Erkenntnis,* XVIII, 223-32.
Thalberg, I. (1964) 'Foreknowledge and Decisions in Advance', *Analysis,* 24, 3, 49-54.
van Inwagen, P. (ed.) (1980) *Time and Cause* (Dordrecht: Reidel).
-- (1982) 'The Incompatibility of Free Will and Determinism', in Watson (1982).
-- (1983) *An Essay on Free Will* (Oxford: Clarendon).
van Straaten, Z. (1980) *Philosophical Subjects: Essays Presented to P. F. Strawson* (Oxford: Clarendon).
Velleman, J. D. (1989) 'Epistemic Freedom', *Pacific Philosophical Quarterly,* 70, 73-97.
--(1992) 'What Happens when Someone Acts?', *Mind,* 101, 2, 461-81.
von Wright, G. H. (1974) *Causality and Determinism* (New York: Columbia University Press).
Waller, B. N. (1990) *Freedom without Responsibility* (Philadelphia: Temple University Press).
Watson, G. (1977) 'Skepticism about Weakness of Will', *Philosophical Review.*
-- (ed.) (1982) *Free Will* (Oxford: Oxford University Press).
--(1982a) 'Free Agency', in Watson (1982).
--(1987a) 'Free Action and Free Will', *Mind,* XCVI, 145-72.
-- (1987b) 'Responsibility and the Limits of Evil', in Schoeman (1987).
Wiggins, D. (1973) 'Towards a Reasonable Libertarianism', in Honderich (1973).
Wittgenstein, L. (1958) *Philosophical Investigations,* trans. G. E. M. Anscombe (Oxford: Blackwell).

Wolf, S. (1987) 'Sanity and the Metaphysics of Responsibility', in Schoeman (1987).
- - (1990) *Freedom within Reason* (Oxford: Oxford University Press).
Zimmerman, M. J. (1984) *An Essay on Human Action* (New York: Peter Lang).

Index