

# Towards a Scientific Conception of Free Will

Mathew Walter Martin Iredale

Ph.D.

University College London

ProQuest Number: 10045699

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10045699

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.  
Microform Edition © ProQuest LLC.

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## Abstract

Can we reach a satisfactory conception of free will that is consistent with the current scientific world view? This important problem is the subject of this thesis.

It is argued that certain areas of science raise profound problems for libertarian conceptions of free will and moral responsibility. Specifically, it is unclear how libertarian free will can arise from the neuronal processes of the brain; it is unclear how, or why, it has evolved; and it is inconsistent with the physicist's conception of time as it is understood through Einstein's Theory of Relativity. These are problems which do not arise for compatibilist conceptions of free will.

Furthermore, there are a number of more philosophical objections which can be raised against libertarianism - including Harry Frankfurt's argument against the principle of alternate possibilities and Galen Strawson's recent objection to ultimate responsibility - which strongly indicate that we do not have free will or moral responsibility in any libertarian sense.

As a consequence of this various compatibilist conceptions of free will are considered in order to see if a satisfactory compatibilist account of free will can be reached. It is concluded that an epistemic account of free will based upon Richard Double's autonomy variable account of free will is the most satisfactory. Several libertarian objections to compatibilism are discussed and rejected.

Finally, some general objections to free will and moral responsibility that have recently been raised by Ted Honderich, Richard Double, and Bruce Waller are discussed. It is argued that they provide no reason to doubt that we have free will. It is therefore concluded that we can reach a satisfactory conception of free will that is consistent with the current scientific world view.

## Table of Contents

1. Introduction	4
2. The Nature of Free Will	18
2.i Personal Experience and Free Will	19
2.ii Libertarianism and Compatibilism	31
3. The U-Condition	43
3.i The U-Condition of Free Will	43
3.ii The U-Condition of Moral Responsibility	71
4. The C-Condition	78
4.i A Problem for the C-Condition	78
4.ii Objections to Frankfurt's Argument	90
4.iii The Theory of Relativity and the C-Condition	103
5. The Evolution of Libertarian Free Will	114
6. Compatibilist Free Will	130
7. Compatibilist Moral Responsibility	151
7.i Objections to Compatibilist Moral Responsibility	154
7.ii Moral Responsibility and Reactive Attitudes	158
8. The Evolution of Compatibilist Free Will	168
9. Objections and Replies	177
9.i Ted Honderich's Objections	177
9.ii Richard Double's Objections	186
9.iii Bruce Waller's Objections	213
10. Conclusions	221

## Chapter 1

### Introduction

I wish to begin this thesis by raising a question: why has the problem of free will lasted so long? In fact, not only has the problem lasted an age, and shown little sign of abating, but if anything it is becoming more complicated. Until quite recently the problem of free will appeared to be quite clear cut, in that nearly everyone agreed that the majority of people have free will: it was just a matter of determining exactly what sort of freedom we had. It makes better copy, perhaps, to highlight the differences between traditional compatibilism and traditional incompatibilism than to highlight their similarities, but there clearly were broad areas of agreement. It was agreed that we had free will, of some sort, and that this enabled us to be morally responsible. And there was almost agreement about which factors remove our freedom; coercion, force, mental illness, and, more recently, hypnotism and brainwashing.

There was of course disagreement about **why** these factors rob us of our freedom, as there still is, but that they did was not in dispute. Occasionally someone would come along, such as Holbach, who would upset this comparatively simple state of affairs by claiming that we were not free, and that both parties - compatibilists and libertarians - had got it wrong. But such people were few

and far between and for the most part no one doubted that we had free will. Now, however, this is not the case. Now there is a veritable barrage of criticism against the view that we have free will (Strawson, 1986; Honderich, 1988b, 1993; Breer, 1989; Double, 1991; and Waller, 1989a, 1989b, 1993 to mention recent critics.) And, just as those who defend free will cannot agree on what sort of free will we have, so those who deny it cannot agree on why we do not have it. But that they deny that we are free cannot be doubted, and their numbers appear to be increasing. Thus not only does the problem of free will look set to continue, it appears to be getting more complex.

So we return again to the question. Why has the free will problem continued for so long, and why, rather than becoming clearer, has it become more complicated? I agree with Honderich (1989b, p.117; 1993, p.105) that to put the lack of progress down to confusion, or unreflectiveness, or word-play, on the part of the protagonists, is quite inadequate. On the other hand, I do not think that his own explanation is fully satisfactory either. He states:

What would explain the persistence of the problem is there not being a single settled idea of freedom, but two ideas, involved in different attitudes. What would go further in explaining the persistence of the problem is each of us having the two ideas, and moving back and forth between them. (1993, p.105)

Whilst I agree that we can have two different sorts of attitudes towards our freedom, I do not believe that we ordinarily do have these two sorts of attitudes. It seems to me that people's ordinary, or common sense, or prephilosophical attitudes are clearly only libertarian, and that compatibilist attitudes are regarded with either

horror or ridicule (for a similar view see Thorp, 1980, pp.34-35.)

Speaking personally, my own prephilosophical, or common sense, views were definitely libertarian (although at the time, of course, I would not have described them as such), and it has been my experience that the initial reaction of nearly every non-philosopher I have met has been to dismiss the idea that free will could be compatible with determinism. This viewpoint, that people are intuitively libertarian, is given support by the fact that the major monotheistic faiths support a libertarian notion of free will, as, I believe, do the majority of legal systems in the world, although not necessarily explicitly. If our intuitive attitudes were not purely libertarian, then I think that it is very unlikely that these two areas would have arisen supporting such a consistently libertarian view of free will.

Contrary to Hobbes, Hume, and others, therefore, I do not think that people ordinarily have a shared view of freedom that is in accordance with compatibilism. There may be aspects of their beliefs which are not inconsistent with compatibilism - that our freedom is threatened if we are forced or coerced - but there are also clearly further aspects which are not in accordance with compatibilism - that we have the ability when we make a decision to decide one thing or another there and then, and that our future is therefore physically open in a way that is quite incompatible with determinism. Rather, people have a shared view of freedom that is in accordance with libertarianism. This is not to say that they **cannot** have

compatibilist attitudes towards freedom, but rather that, ordinarily, they do not.

Why, then, has the free will problem persisted for so long? My own view is that the free will problem has persisted because our ordinary view of freedom is at odds with the evidence. To be more specific: we intuitively interpret our experiences of freedom in a libertarian way, and this acts as a natural barrier to our acceptance of the view that we do not have libertarian free will. However, when we look more deeply into the subject, it does not appear as if the conditions necessary for libertarian free will can be fulfilled. This then leads us towards a compatibilist or hard determinist solution, which, to come full circle, appears to be at odds with our experiences of freedom.

It does not appear as if the conditions necessary for libertarian free will can be fulfilled because of a number of problems which are associated with those conditions. For example, it seems to be impossible to provide a satisfactory, that is, detailed and coherent, analysis of the conditions of libertarian free will - as a number of libertarians have admitted (Wiggins, 1973; Thorp, 1980; van Inwagen, 1983). Furthermore, in as much as they can be clarified, the conditions of libertarian free will consistently appear to be at odds with the prevailing scientific world view (whether they actually are or not is another matter, but they consistently appear to be.) They are inconsistent with the deterministic world of classical physics, and appear to gain no great advantage from the indeterministic world of quantum physics. Add to this the



problems that can be raised in connection with the neurosciences (Honderich, 1988a), the Theory of Relativity (discussed in chapter 4 of this thesis) and evolution (chapter 5), and it becomes clear that the libertarian has some considerable work to do if he is to provide a scientifically acceptable account of free will.

In short, we are torn between two positions, one which accords with what we feel, and one which accords with what we learn about the world. Thus although I agree with Honderich that there is no single settled idea of freedom, but two ideas (a view which will be defended in chapter 2,) I disagree with his assertion that we are torn between these two ideas because we move back and forth between two sets of attitudes. Rather it is that we are drawn to one by our intuitions and drawn to another by other evidence.

Furthermore, this explanation of the persistence of the free will problem, unlike Honderich's, explains why the free will problem only arises for those who study free will, and why it does not arise for everyone else. For as far as the majority of people are concerned there is no free will problem: their beliefs, or attitudes, concerning free will remain quite untroubled by the everyday events of life, which appear to be perfectly consistent with such beliefs/attitudes. It is only those who take a deeper interest in free will who discover that this simple state of affairs is not the case, and that there are a number of problems with the concept of free will.

This situation is not something that Honderich's explanation can explain. For if the persistence of the

problem is due to our not having a single settled idea of freedom, but two ideas, involved in different attitudes, then why do the philosophical problems associated with free will not trouble everyone, as opposed to just philosophers? If each of us have these two ideas of freedom, and move back and forth between them, then why is there such an apparent lack of concern amongst ordinary people as to the problems which this would surely raise? Surely the problems raised by this curious state of affairs would be everyone's concern if it were the case? They would argue about what sort of attitudes one should rightly hold concerning free will in the same way that they argue about what sort of political or religious attitudes or ideas one should rightly hold. But clearly they do not. As Dennett says, echoing Dewey, people worry about a lot of things but metaphysical freedom is not one of them (1984, p.5.) This is because they do not have these dual attitudes regarding free will, but only one, libertarian, attitude.

A further contributory factor to the longevity of the debate, and one reason for its increasing complexity, has arisen because of the exasperated feeling of some philosophers that if either one of these two 'so overridden and wearied nags' - compatibilism or libertarianism - was really right then one of them would have proved it by now and plodded 'at last into the winner's enclosure' (Honderich, 1988b, p.3). The fact that neither nag clearly has beaten the other into the winner's enclosure causes these philosophers to conclude that they are quite possibly both wrong, and to seek an alternative

solution to the problem - that we do not have free will.

Honderich is not alone in despairing at the way that the same two combatants have parried back and forth for so long. Consider the following passage from Richard Double:

I have come to the non-realist position grudgingly. Several years ago at a meeting of the American Philosophical Association a very prominent incompatibilist commentated on a paper delivered by a younger, less prominent, but very sharp compatibilist. The exchange between the two lasted the entire hour, and towards the end it became clear that neither speaker could understand at all why the other held the position that he did: as one spoke, the other just shook his head in disbelief. I remember that I was surprised that these two very bright people could find so little of value in what the other said, and it seemed implausible to me that their predicament could be explained away by way of disagreements over terminology, logical blunders, or differing estimates of empirical probabilities. Instead, I thought, something deeper had to be at stake. (1991, pp.5-6.)

Whilst I too must admit to being occasionally exasperated by the way that the same old arguments are trotted out time after time, I do not (yet) think that the time has come to abandon free will. I believe that we are nearer to a solution than opponents of free will believe, or libertarians would like to admit to. Hopefully, the arguments of the thesis will bear me out in this belief.

I began the thesis with a consideration of why the free will debate has lasted so long because I think that the answer to this question will enable us to answer the problem of whether or not we have free will, and, if we do, just what sort of freedom it is. I stated that the longevity of the problem is due to the fact that whilst we are lead by our intuitions to believe that we have libertarian free will, such a conception of free will appears to be subject to a number of fatal objections. If this is an accurate explanation then there would seem to

be two clear tasks that need to be tackled before any others if we are to resolve the problem. First, it needs to be determined whether or not our intuitive interpretation of our experiences of freedom - that we have libertarian free will - is the right interpretation. It may be that these experiences are also consistent with a compatibilist conception of free will. Secondly, it needs to be determined if the various philosophical and scientific arguments raised against libertarianism do in fact undermine it, or if they merely appear to because of inadequate or fallacious reasoning. In other words: can we actually fulfil the various conditions necessary for libertarian free will?

But tackling these two tasks alone will not necessarily resolve the problem. Although I have stated that the problem of free will arises because of a conflict between the direction in which we are lead by our experiences of freedom and the direction in which we are lead by various philosophical and scientific arguments, this is somewhat of a simplification. This really only explains why the conflict arose in the first place, and why it continues to arise in the same form between subsequent generations. In order to understand why the conflict is so persistent we must appreciate that the libertarian does not just object to compatibilism because of his intuitions, but because he believes that compatibilism suffers from a number of inherent faults (although his intuitions are, I think, the primary motivation behind these further objections.) These objections include, depending upon the critic in question

and in no particular order of importance: an inability to explain how we can be rational if we are determined; an inability to explain how we can have real alternatives open to us, or how we can meaningfully deliberate, if our decisions are determined; and an inability to explain how we can justifiably be held morally responsible if our actions are determined.

It is these issues, as much as his intuitions, that cause the libertarian to find compatibilism unacceptable, and which must be considered by the compatibilist if he is to provide a satisfactory account of free will. Of course, these issues have been considered many times before, and in some cases I think that the answers offered have been very satisfactory: most obviously, with regard to the problem of whether we can be held to be rational agents if determinism is true. Despite the fact that some libertarians continue to raise this objection (most recently Lucas, 1993), I fail to find it at all persuasive. It has recently been discussed, and rejected, by compatibilists (Dennett, 1984), by libertarians (Swinburne, 1986), and by those who reject both these positions (Honderich, 1988a; Double, 1991.) Since I do not believe that I can add anything of any substance to these discussions I do not propose to discuss this particular problem any further. As to the other problems, although they have also been discussed in some detail such discussions have, to my mind, been less satisfactory, and so I shall have more to say about them later in the thesis.

There are, then, three main issues that need to be

tackled in the free will debate. Firstly, it must be determined whether or not our personal experiences of freedom point us towards any particular conception of free will; secondly, given that they at least support libertarian free will it must be determined whether or not the conditions necessary for libertarian free will can actually be fulfilled; and thirdly, if they cannot be fulfilled, and if our personal experiences of freedom also support compatibilism, it must be determined whether or not a compatibilist conception of free will can be expounded which overcomes the various objections raised above. It is these issues which I discuss in the remainder of the thesis.

In chapter 2 I discuss our experiences of freedom in order to determine whether they support any particular conception of free will. I argue that our experiences of freedom come from two areas: our observations of other people's behaviour, and our personal experiences of deliberation and decision making. I conclude that although we intuitively interpret these experiences in a libertarian way, they are also quite consistent with a compatibilist conception of free will (contrary to the beliefs of certain libertarians, such as van Inwagen, 1983.) Having established that our experiences are consistent with these two quite different conceptions of free will, I then consider what conditions we would have to meet in order to be free in both a libertarian and a compatibilist sense. This then sets up the rest of the thesis: which of these conditions, if any, can we fulfil?

Since people are invariably lead by their experiences

towards libertarian free will, it is to this conception of free will which I turn first. In chapters 3 and 4 I consider whether we can fulfil the two main conditions of libertarian free will, which (after Klein, 1990) I term the U-condition and the C-condition.

In chapter 3 I argue that the U-condition as it applies to free will cannot be fulfilled unless we wish to reject the current scientific world view, and that as it applies to moral responsibility it is logically impossible.

In chapter 4 I argue that the C-condition is demonstrably superfluous to ascriptions of free will, and, even if it were not, that it is clearly inconsistent with Einstein's theory of relativity. I conclude that it does not appear as if we can fulfil the conditions of libertarian free will.

In chapter 5, recognising that my arguments of the previous two chapters will not be considered to be decisive by many libertarians, I consider one further objection to libertarianism that has rarely, if ever, been seriously raised. It concerns the evolution of libertarian free will. I argue that it is quite unclear how libertarian free will could have evolved through the small successive changes that evolution demands. Furthermore, it is equally unclear why libertarian free will should have evolved: it appears to confer no advantage upon those agents that are said to possess it compared to those that do not. I conclude at this point that libertarian free will, although it has the undeniable support of our intuitions, is otherwise quite unsatisfactory. It provides

no solution to our problem.

Consequently, in chapter 6, I turn to compatibilism. I discuss several different compatibilist accounts of free will and conclude that an epistemic account is the most satisfactory.

In chapter 7 I discuss and dismiss various objections that have been raised against compatibilism concerning moral responsibility.

In chapter 8 I turn to the evolution of compatibilist free will. I consider why it has evolved and why it offers a clear advantage to those individuals that possess it compared to those that do not. I conclude at this point that compatibilism provides the most satisfactory solution to the problem of free will.

In chapter 9 I discuss a number of recent objections to the general notions of free will and moral responsibility by Ted Honderich, Richard Double, and Bruce Waller.

One final point before I continue. Although there are many issues in the free will debate which are of importance, I believe that one of them is of greater significance than the others. This is the issue of whether free will is consistent with the current scientific world view. If there is a theme running through this thesis that must be emphasised it is that the only acceptable solution to the free will problem must be one that is consistent with the current scientific world view. Hence the title. Traditionally, the free will debate has centred exclusively around determinism, and, with a few recent



exceptions, this continues to be the case, and even amongst the exceptions I think that few, if any, have explicitly called for it to be centred around science. But it seems clear to me that this is unsatisfactory. The issue of whether we have free will or not given determinism should be explicitly replaced by the issue of whether we have free will or not given the current scientific world view. This is not simply because modern physics has displaced the deterministic world view of classical physics, but because there are a whole range of scientific ideas - concerning evolution, the neurosciences, the nature of time - which are directly relevant to free will and which are quite separate from the issue of whether the universe is fundamentally deterministic or not.

Science offers us a way of judging between different conceptions of freedom without begging the question in favour of one conception of free will over another, or without begging the question in favour of the view that we have free will in some meaningful sense over the view that we do not. This is because science has no philosophical axe to grind - at least, not with respect to free will - and so offers us a clear way in which to resolve the free will problem.

Of course, in taking this approach I am assuming that science is taken to be an acceptable way in which to judge between different conceptions of free will, but I know already that this will by no means please everyone. I can think of at least two libertarians who would rather compromise the consistency of their conception of free

will with a scientific world view than compromise their views of free will, and there are probably more, but I do not believe this to be a rational course of action. To stick to a view in the face of science - such as the religious creationists do in America - smacks to me of irrational desperation. Obviously many scientific theories are not proven fact - to echo Popper, they are only conjectures - and so they could, in theory, be wrong. But I think that when one is faced by a theory that has a wealth of empirical evidence to support it, such as evolution, and when there are no serious alternatives, to object to it because it conflicts with your own personal beliefs or intuitive feelings, is simply unjustifiable.

Having said that, I do wish to stress that this thesis will be no piece of 'scientism' (to use van Inwagen's phrase, 1983, p.215) That is, I have no intention to disparage all that is not science. Rather, the thesis will involve a range of arguments of which the scientific ones are just one, albeit significant, part. However, it should become clear during the course of the thesis that the weight of non-scientific evidence serves to reinforce this line of approach rather than undermine it. Science only serves to light a path which is already there.

## Chapter Two

### The Nature of Free Will

In chapter one I stated that the first task of any inquiry into free will is to determine whether or not our ordinary, everyday experiences of freedom point us towards any particular conception of free will. This will be undertaken in the first section of this chapter. I shall argue that our experiences are consistent with both libertarian and compatibilist conceptions of free will. This clearly goes against the view put forward by some libertarians that our experiences are inconsistent with determinism (e.g. van Inwagen, 1983), but it has something in common with several recent discussions which also reject this libertarian argument (Strawson, 1986; Honderich, 1988b, 1993; Double, 1991.) It also differs in certain respects, though, as has already been mentioned in chapter 1, and as I shall discuss more fully in chapter 9.

Having shown that both compatibilism and libertarianism fulfil one important condition of free will - consistency with our experiences of freedom - I shall then go on to discuss two other conditions which must be satisfied by a given account of free will if it is to be considered as a potential solution to the free will problem. It must be empirically plausible, and it must explain how we can be morally responsible.

## 2.i Personal Experience and Free Will

Our personal experiences of freedom can arise from two areas - from our observation of others and from our experience of our own deliberation and decision making. I shall consider each of these areas in turn, beginning by discussing the conclusions we can draw from our observations of others.

Our observations of other people begin almost from birth. We learn at an early age to distinguish between certain individuals (who, as adults, we have come to regard as free) and certain other individuals (who we have come to regard as not free) purely on the basis of the behaviour of those individuals, and the behaviour shown towards them by others. We need not, and in most cases do not, make the distinctions we make because we believe certain individuals to be free and certain others to be unfree. We need not have any concept of free will at all. We simply see that certain people lack abilities which others have, and that this requires a modification in behaviour.

As we grow older and our beliefs become more sophisticated we learn that there are certain individuals who are generally held by society to be capable of free decisions, to be potentially free individuals, and certain individuals who are not. We learn that normal, adult, humans are taken to fall into this first category, whilst animals, and those who are mentally ill or brain damaged, are taken to fall into the second category. Such individuals are taken to be simply incapable of free

decisions. At the same time we learn that there are certain situations which reduce, or completely remove, the capability for free decisions which potentially free individuals have. Such situations include brain damage, hypnotism, brainwashing, forceful threats or pressures, and situations, if the individual is phobic, manic or neurotic, which involve the individual's phobia, mania or neurosis.

Individuals may disagree about where exactly to draw the line between freedom and unfreedom, but for the moment I shall assume that the distinctions outlined above are those that are generally held by society as a whole. They are held by the majority of individuals who make up our society, and are an integral part of the laws and religions that guide our society.

I now wish to consider what grounds we have for making these distinctions. What reasons can we find, beyond our desire to conform to society's expectations, for holding such beliefs about the nature of people's thoughts? Why do we believe that some people are capable of free decisions whereas others are not?

Since we cannot have direct access to a person's thoughts the only way in which we can learn about those thoughts is through their behaviour: from what they say and what they do. Therefore if we wish to know how to distinguish between free and unfree decisions we must determine what it is that distinguishes the behaviour that accompanies such decisions. We must determine what it is about certain people's behaviour that leads us to believe that they are capable of making free decisions and what it

is about other people's behaviour that leads us to believe that they are not capable of making free decisions. We must determine what, if any, are the common features of the behaviour of individuals who we regard as free that enable us to identify them as free and what, if any, are the common features of the behaviour of individuals who we regard as unfree that enable us to identify them as unfree. Having identified the common features of free and unfree behaviour we can then seek an explanation for this behaviour.

If we compare those individuals who we regard to be free with those who we regard to be unfree then we can see that the one feature that divides the two is self-control. Free agents are in control of their behaviour, whereas unfree agents are not.

Some unfree agents, for example, are clearly under the control of another agent: hypnotised agents are under the control of the hypnotist; brainwashed people, less directly, but just as completely, are under the control of those who brainwashed them. Others behave in such a way as to lead us to believe that they have little or no control over their thoughts even though there appears to be no agent controlling them (e.g. mentally handicapped and psychotic subjects.) Consider, for example, R. E. Kendell's description of a typical schizophrenic:

The subject ceases to experience his mental processes and his will as under his own control; he may insist that thoughts are being put into his mind or removed from it by some alien force, or suspect that he is being hypnotised. He hears voices telling him what to do, commenting on or repeating his thoughts, discussing him between themselves, or threatening to kill him. In the acute stages of the illness other hallucinations and delusions of varied kinds may be present. His thought processes also develop a characteristic vagueness and illogicality. At first he

simply keeps wandering off the point, but in some chronic patients there eventually ceases to be any logical connection between one idea, or phrase, and the next; so their speech becomes almost incomprehensible. (1987, p.697)

In such a case, although we clearly do not observe the agent to be under the control of another, we still believe that he is not in control of his decisions. We might describe the situation as one in which the agent is under the control of his psychosis, or we may prefer to state that his behaviour is limited by his psychosis and that there is nothing in actual control of his behaviour (if, like Dennett, 1984, p.52, we believe that only agents can control things,) but either way it appears that the agent has no alternative but to behave in the way that his psychosis causes him to behave. He cannot choose to behave 'unpsychotically' as it were, even if he wants to. He is powerless; he cannot help what he does, and so cannot be said to be in control of his decisions.

Loss of control occurs, therefore, if the agent has no alternatives open to him. But this is not the sole cause of loss of control, for there are certain agents who we believe have alternatives open to them, but who we would not wish to regard as in control of their behaviour (e.g. certain mentally retarded agents and those under such great mental strain that we say that they were 'not themselves' when they acted.).

In such cases it is not the absence of alternatives that robs the agent of control, for the agent is not compelled to make the decisions he makes, either by another agent, or by some psychosis, but rather it is the agent's inability to choose between the alternatives that

are available to him in a rational or non-wanton manner. The agent appears to be incapable of the rational deliberation that free agents are capable of when making choices (even if free agents do not always actually deliberate.) This is not to say that such agents are incapable of making rational or non-wanton decisions, but rather that if they do then this will merely be a happy accident rather than due to any conscious effort on their part. (By wanton decisions I mean decisions which are not made with any clear purpose in mind. They are therefore to be distinguished both from irrational decisions - or decisions which appear to be irrational by others' standards - and from spontaneous decisions, both of which may be free. For example, if one Sunday afternoon I suddenly decide to go to the cinema, although the decision is spontaneous, and, given my poor financial situation, somewhat irrational, it can still, I think, be a free decision, since it was my decision, and I had a clear alternative open to me: to not go to the cinema. How irrational a decision may be before it can no longer be considered to count as free is a moot point, and I shall consider it further in chapter 6.)

Consequently, it can be concluded from our observations of the behaviour of others that the one feature which free agents have in common, and which unfree agents lack, is self-control. Agents may lack self-control either because they have no alternatives available to them, or, if they do have alternatives available to them, because they are unable to judge between these alternatives in a rational or non-wanton manner.



What we do not observe in the behaviour of others is a lack of determinism, or the presence of contra-causal freedom, agent causality, or any other metaphysically curious phenomenon. Nor, it must be said, do we observe anything which denies the presence of such phenomena. In other words, there is nothing in our ordinary observations of each other to allow us to decide between compatibilist or libertarian conceptions of free will, since both are consistent with the view that freedom of the will can be described in terms of self-control. (Of course, libertarians and compatibilists define self-control in different ways, as we shall see in the next section, but this is, for the moment, unimportant, since both definitions are consistent with the observations concerning self-control made above.)

This basic observation - that free will requires some degree of self-control - is also supported by our personal experiences of freedom. Under normal circumstances, in most situations, we really feel as if we are in control of our decisions, and not under the control of another. We feel as if we have alternatives open to us, and that we are able to choose between them in a rational or considered way; our choices are not foisted upon us or chosen wantonly. Galen Strawson provides a useful example to illustrate this feeling:

You set off for a shop on Christmas Eve, just before closing time, intending to buy a bottle of whiskey with your last ten-pound note. On the steps of the shop someone is shaking an Oxfam tin. You stop, and it seems very clear to you that it is entirely up to you what you do next - that you are truly, radically free to choose, in such a way that you are ultimately morally responsible for what you choose. (1986, p.viii)

Strawson's example also forcefully brings out the feeling that we often have that when we make a decision we have it in our power, there and then, to make a different decision. We feel that many of the decisions we make are not ones that we have to make. We really feel as if we have alternatives open to us, there and then, that in the exact situation that we find ourselves we can choose one thing or another.

It is this feeling which causes some libertarians to claim that our personal experiences of freedom are incompatible with a compatibilist conception of free will. They argue that we can only make sense of this personal experience of freedom - the feeling that we can, there and then, either decide one thing or the other - if the future is physically open. For, they reason, if the future is physically closed, which it is if we are determined, then we clearly cannot, there and then, either decide one thing or the other, and such an experience of freedom would simply be pointless.

I admit that it does appear that we have good grounds to believe that we have libertarian free will because of this personal experience. It is a persuasive argument. For if we feel like we can choose between two physical alternatives then the most obvious explanation for this feeling is that we really can so choose. On the other hand, it does not automatically follow from either the persuasiveness or the obviousness of this argument that we can **only** make sense of our feeling of freedom in terms of libertarian free will. It may be that we have grounds, from our personal experiences of freedom, to believe that

we have libertarian free will, but I also believe that we have grounds, from these same experiences, for believing that we could have compatibilist free will.

The point at issue is whether this feeling of freedom really commits us to the view that the future is physically open. If the compatibilist denies this, if he wishes to assert that we are not committed to the view that the future is physically open, then he has to explain why we have this feeling of freedom. For surely there would be no reason for such a feeling if we did not actually have the sort of freedom it implies?

Although this argument appears to be sound it is really quite fallacious. It proceeds according to the following premise; that there is only one possible interpretation of our personal feelings of freedom. But why should we assume this to be true?

Consider the medieval interpretation of the sun moving across the sky. It was generally held to be evidence for the fact that the sun must revolve around the earth. But once one knows that the earth rotates around the sun one can see that the movement of the sun across the sky is also consistent with this situation. What we have, then, is a situation in which a state of affairs appears to support one set of facts, but in actual fact also supports quite another. In other words, there are two possible interpretations of the observation that the sun moves across the sky. The obvious one, that the sun is moving and that the earth is still, and the correct one, that the earth rotates around the sun.

Exactly this situation exists with regard to our

feelings of freedom. There are two possible interpretations. The obvious one, that we have libertarian freedom, and that the future really is physically open, and the compatibilist one, that this feeling is a necessary part of rational decision making, quite regardless of whether the future is physically open or not.

It is a necessary part of rational decision making even in a universe with a physically closed future because the future is unknown. We do not know in many situations what the future will hold, although we usually have a pretty good idea, and so in order to make the best decisions we have to consider the possibility that any number of things could happen. Not because they really could all happen, but simply because we do not know which one of them will happen. In other words, deliberation does not require that we have physical alternatives, but only epistemic ones.

We deliberate over two or more courses of action in order to determine the best course of action that we should take - that is, the most desirable course of action we should take - in order to achieve some goal. We consider a range of options, any of which, as far as we know, we could be determined to take, and then opt for the one which we consider to be the best. According to the libertarian such deliberation only makes sense if we assume that the future is physically open - that all the options we consider are physically possible for us. But the compatibilist realises that deliberation also makes sense even if we assume that the future is physically

closed, but unknown. For it does not matter if the decisions we do make are the only ones we can make, so long as the decisions that we make are the best ones to make. And the only way we can determine if the decisions we make are the best ones to make is by considering all the decisions that, as far as we know, could be carried out to achieve our goal.

The situation, for the compatibilist, is something like this: we are determined, through evolution, to make what we consider is the best possible choice we can in order to achieve some goal, G. What we consider to be the best possible choice may be highly subjective, and even at times irrational by others' standards, but for us it is the best choice to make. We cannot avoid making this choice. But until we consider all of the possible ways in which, for all we know, we could achieve G, until we deliberate, we cannot know what this best choice is. Suppose that we believe that we could achieve G by either choosing A, B, or C. In order to determine which of these choices is the best one to make we have to consider the desirability of each choice. We have to deliberate over which will best enable us to achieve G. We may, for various reasons, decide that B is the most desirable, the most likely to enable us to achieve G, and so we then choose B. And we may accept that we were determined to choose B, but we could not have chosen B, and be confident that it was the best choice to make, without deliberating about it.

The libertarian may reply that we could not help but choose B, since we were determined to choose it, and that

we therefore had no genuine alternatives. But this simply begs the question: why should the only genuine alternatives be physical ones, when epistemic alternatives, as much as physical alternatives, enable the agent, through deliberation, to make the best possible choices to achieve his aims?

In fact, there are valid objections that can be raised against the concept of epistemic alternatives, but I shall not pursue these objections here (I consider them in chapter 6.) This is because for the moment I only wish to show that our personal experience of freedom is compatible with a physically closed future, if that future is unknown, and such objections do not affect this conclusion.

There is one further reason, I suggest, for the belief people have that their personal experiences of freedom provide evidence for the fact that they have libertarian freedom. Such people will have observed, probably many times, someone who was faced with a choice which was very similar to one which they themselves have been faced with and who made a different choice to the one which they made when they were faced with that choice. They therefore reason that since this person was able to make a different choice to the one that they made when they had to make the choice that this provides some sort of evidence that they too could have made a different choice. But this argument assumes that the choice faced by the other person is exactly the same as the choice faced by the first person, and this is clearly not true. Therefore any conclusions based purely upon this

assumption will also be untrue.

Furthermore, in a determined world we would, of course, expect to see people making different decisions to ones we made in similar situations. Therefore such observations no more provide evidence of libertarian free will than does the fact that two trees can grow in the same area, with the same light, soil, etc, and yet grow in quite different ways, provide evidence of free will in trees.

Note also that there is absolutely no other evidence to back up the libertarian's interpretation of our personal feelings of freedom. We are simply unable to go back in time in the manner required to prove that freedom of this sort exists. Our observation of the behaviour of free individuals only enables us to infer that they can change their behaviour on subsequent similar situations, because that is all that we observe. We never observe someone in exactly the same situation, at the same place and time, making a different decision to the first one that they made. And this is the only observation which could lead us to conclude that we have this sort of freedom, and that the future is really open.

This concludes my discussion of our experiences of freedom. It should now be clear that there are two possible interpretations of these experiences. The first, the one that people most readily jump to without further thought, is that we must have libertarian free will; that the future must be physically open. The second, the one towards which I am personally most inclined, is that such experiences do not require that the future is physically

open, and that they are therefore consistent with compatibilist free will.

## 2.ii Libertarianism and Compatibilism

What conditions do we have to fulfil if we are to have the sort of free will that our experiences lead us to believe that we have? What specific conditions have to be met by libertarian and compatibilist conceptions of free will if they are to be consistent with our experiences?

We have seen that our experiences of freedom lead to the conclusion that to be free is to be in control of one's decisions. And to be in control of one's decisions one needs to have alternatives open to one and to be able to choose between them in a rational or at least non-wanton way.

This leads to the formulation of a general condition of free will that is common to both compatibilist and libertarian conceptions of free will: indeed, a condition which must be met by any conception of free will if it is to be considered as such. Here is the condition:

To be considered as satisfactory a given conception of free will must be able to explain in a coherent and unambiguous way how we can be in control of our decisions - how we have alternatives open to us and how we are able to choose between them in a non-wanton way - in a way that accords with our experiences. That is, it must enable us to distinguish, and to justify the distinction, between those who we ordinarily take to be in control and those who we do not.



However, by itself this condition will not be sufficient to enable us to reach a satisfactory solution to the problem of free will. For we may be provided with a coherent and unambiguous account of self-control that we are simply unable to possess. We therefore require the following additional condition:

To be considered as satisfactory a given conception of free will must be empirically plausible. That is, not only must it fit in with our various experiences of freedom but it must also fit in with our empirical view of the world. The more a conception deviates from this view, the less satisfactory it is.

Richard Double is one of the few philosophers to have explicitly stated that this is a necessary condition of any account of free will. He states that we should ask the following of any account of free will:

Does the account need a one-in-a-million shot to be satisfied? To take a simple example, if one held that dreams are real events occurring to people while travelling in 'astral bodies'...the empirical plausibility of that account would seriously count against it. (1991, p.218)

It may be thought that by including this condition I am automatically prejudicing the thesis right from the outset against certain libertarian conceptions of free will. This may turn out to be the case, but I do not believe that this should cause us to exclude an empirical condition. I do not think that it is asking too much of libertarian (or indeed compatibilist) accounts of free will that they accord as closely as possible with our empirical view of the world, especially when an increasing number of libertarians are clearly aware of the need to

make their accounts of free will empirically plausible (Thorp, 1980; Kane, 1985.) And consider the alternative: if we did not include an empirical condition then any coherent account of free will that accorded with our experiences, no matter how empirically implausible, would have to be considered as acceptable, and this is surely not a satisfactory state of affairs?

There is one final condition that needs to be set out. It concerns the common sense belief that free will enables us to be morally responsible for our actions. To be satisfactory, a given conception of free will must satisfy this deeply held belief. Such a view has been expressed by philosophers of very different sentiments:

Without free will, we should never be morally responsible for anything. (van Inwagen, 1983, p.209)

The varieties of free will we deem worth wanting are those - if there are any - that will secure for us our dignity and responsibility. (Dennett, 1984, p.153)

It would not be too strong to say that, for many philosophers, the whole point of trying to explicate the **free will** concept is to show how it is possible for persons to be morally responsible. (Double, 1991, p.75)

If Double is right on this last point then in my opinion such philosophers have their priorities wrong. It is one thing to be aware of the relevance of the issue of moral responsibility to the free will debate, but it is quite another to place that issue exclusively at the centre of the debate, for then one places too much emphasis upon the issue of moral responsibility at the expense of the other issues (as Honderich has recently discussed, 1988b, p.11). However, at the same time it cannot be denied that the issue of moral responsibility does have a special place in the free will debate, and one

would be equally mistaken in one's priorities if one did not explain how free will enables us to explain and justify our practice of holding ourselves and others morally responsible.

These last two conditions are not that precise, but necessarily so, since to be more specific would beg the question and therefore might possibly exclude certain conceptions which would more profitably have been included. There is a further condition which some may want to include. It is that any conception of free will should be, in the words of Daniel Dennett (1984), worth wanting. It should be something the lack of which would be undesirable. It should give our lives a certain value or meaning. It should not be something towards which we have no feelings, or only impartial ones. I have no great objections to including this amongst the general conditions, although I think that any conception that fulfils all the other conditions will automatically be worth wanting. It therefore seems somewhat superfluous to specify it as a condition in its own right.

Of these three conditions, detailed attention has really only been paid to the first and the third in the free will literature. In fact, many discussions of free will, consciously or unconsciously, have virtually ignored the issue of empirical plausibility. I believe that this is a significant omission, and one which this thesis is conscious of avoiding. For the remainder of this chapter, though, I shall only be concerned with the first condition, self-control.

I have argued that our experiences of freedom can

lead us to view self-control in two distinct ways, and I now wish to consider them in more detail. One way, that of the libertarian, believes that we cannot be in control of our decisions if we are determined to make the decisions we make. The other way, that of the compatibilist, believes that we can be in control of our decisions even if we are determined to make the decisions we make.

The libertarian believes that we cannot be in control of our decisions if we are determined for two reasons. Firstly, he believes that if our decisions are determined that they cannot really be **our** decisions. They will simply be 'the consequences of the laws of nature and events in the remote past.' (van Inwagen, 1983, p.56) Secondly, he believes that the sort of alternatives that we require in order to be in control of our decisions are physical alternatives, and physical alternatives are not possible in a deterministic universe.

This has lead the libertarian to formulate two specific conditions for free will. The first states that our decisions, in order to be ours, need to have originated within us. They are not connected by a causal chain to events which originated outside us. They are not due only to our heredity and the action of the environment. This condition is important to the libertarian because he defines moral responsibility in terms of being ultimately responsible for one's acts, and the decisions which lead to those acts. If our decisions do not ultimately originate within us then we cannot be ultimately responsible for them, and so we cannot be morally responsible. C.A.Campbell provides a succinct

description of this condition when he states that a free decision 'must be one of which the person being judged can be regarded as the **sole** author.' (Campbell, 1957, p.160) He clearly realises that such a capacity is inconsistent with our being determined solely by our heredity and the environment:

..if we are mindful of the influences exerted by heredity and environment, we may well feel some doubt whether there is any act of will at all of which one can truly say that the self is sole author, sole determinant. (1957, p.160)

More recently, Randolph Clarke has stated that free will requires what he terms the condition of production (CP):

When CP is fulfilled, an agent is a real origin of her action. She determines that she perform that action, and that determination by her is not determined by anything beyond her control.

Any account of free will that allows that all events (except perhaps the world's first event) are caused, that all causes are events, and that all causal chains go back in time, if not forever, then to the beginning of the universe will fail to secure CP, regardless of whether causal relations are deterministic or probabilistic. (1993, p.193)

The first condition of libertarian free will is therefore that the agent's decisions ultimately originate within the agent. In the remainder of the thesis I shall refer to this condition of free will as the U-condition (after Klein, 1990.)

The second condition requires that we have physical alternatives open to us (and that we are able to choose between them in a rational or non-wanton way.) We have the ability to decide one thing or another, there and then, in the exact circumstances that we find ourselves. The traditional way in which this condition is put is to state that we have the ability to choose otherwise under the

exact conditions that obtain at the moment of choice. John Thorp, for example, states that 'A decision is free if the agent could have decided otherwise' (Thorp, 1980, p.7), whilst Jennifer Trusted puts it more explicitly:

When libertarians say that there is free choice, they are saying that an action is free because the agent could have chosen to act differently and could in fact have acted differently. He or she could have chosen and acted differently despite all the physical conditions remaining the same. (Trusted, 1984, p.94)

After Klein I shall refer to this condition as the C-condition, the 'could have decided otherwise condition', after the libertarian's claim that when he made a free decision he could have decided otherwise.

These are the two basic conditions which libertarians state must be met by an agent if that agent is to be considered to be in control of his decisions, and therefore to be free. Compatibilists, of course, do not believe that either of these conditions must be met in order for us to be in control of our decisions in a way that is consistent with our experiences.

The compatibilist denies the libertarian assertion that an agent must be the ultimate originator of his decisions in order for those decisions to be really his. This is because he is not concerned with defending the mysterious notion of ultimate responsibility, which requires that the agent be the ultimate originator of his decisions. As far as he is concerned it is sufficient that our decisions arise within us in order for them to be really ours. If they arise from the processes of our brains then they are our decisions. It is as simple as that. Compatibilism does not need a more complicated

definition.

I must stress here that when I say that a decision is ours, I am not stating that it is ours because we are in control of it, but only because it arises within us. I mention this to avoid misunderstanding. People often say such things as 'it was my decision', meaning not just that the decision arose within them, but also that they were responsible for it, that it was a decision taken under their control. This somewhat loose sense of 'my decision' - really meaning 'my decision, for which I also take responsibility' - should not be confused with my use of the phrase 'my decision', which is purely intended to convey the fact that the decision arises within the agent in question, and says nothing about whether it was under his control or not. All sorts of decisions may arise in agents, and therefore be said to be their decisions, even though they are not ones which they have any control over (because they are caused by psychoses, brain damage, hypnosis.)

Consequently, pace the libertarian, simply because I am determined to make a decision it does not follow that that decision is not mine. If it originally arose within me then it is as much mine as it could ever be. The libertarian's notion of origination cannot improve upon this state of affairs. If we have two agents - one, O, who can originate his decisions, and one, A, whose decisions arise within him - then in what way is O's decision more his own decision than A's decision? Remember that I am not presently concerned with control, with who has the greater control of his decisions, but merely with 'ownership'. In

what way are O's decisions more his own than A's decisions are A's? I cannot think of any answer that does not beg the question (e.g. they are not A's decisions because he did not ultimately originate them.)

To turn to alternatives, as far as the compatibilist is concerned not only do we not have physical alternatives open to us of the sort described by the libertarian, we do not require such alternatives in order to exercise free will.

If determinism is true, then clearly, in some sense, there are no alternative possibilities. Relative to the laws of nature and antecedent conditions, it is not possible that one does anything but what one does. Compatibilists must dismiss this sense as irrelevant to free agency. (Watson, 1987a, p.154)

According to Watson, and many other compatibilists, the alternatives that we require for free will are what can be termed conditional alternatives. They replace the libertarian's phrase 'he could have decided otherwise' with 'he would have decided otherwise, if he had wanted to, or if conditions had been slightly different.' Kenny, for example, states: 'The type of power to do otherwise which is necessary for freedom is the power to do otherwise if one wants to.' (1978, p.26) More recently, Thornton has stated:

..saying that we can choose otherwise than we do is compatible with saying that we would choose otherwise only if there were some different causal factor operating; and saying that we would choose otherwise only if there were some different causal factor operating is compatible with determinism. (1989, p.131)

They are conditional alternatives since they depend upon certain counterfactual conditionals being true (unlike the libertarian notion of alternatives which is categorical.)



However, contrary to Watson's assertion (1987a, p.154), the compatibilist does not need to defend a conditional account of alternatives (c.f. Wolf, 1990, pp.97-100.) Recall the discussion of deliberation in the last section. I stated that the compatibilist holds that an agent is determined to make what he consider is the best possible choice he can in order to achieve some goal, G. However, he also emphasises the fact that until the agent considers all of the possible ways in which, for all he knows, he could achieve G, until he deliberates, he cannot know what this best choice is. If he believes that he could achieve G by either choosing A, B, or C, then in order to determine which of these choices is the best one to make he has to consider the desirability of each choice. He has to deliberate over which will best enable him to achieve G. Now suppose that he decides that B is the most desirable, the most likely to enable him to achieve G, and so he chooses B. According to the conditional compatibilist's version of events, although the agent may accept that he was determined to choose B, and that from one point of view he had no alternatives open to him, he can also assert that from another point of view he did have alternatives open to him. This is because it is also true that if he had decided through his deliberation that A was the best choice to make, he would have chosen A, and if he had decided that C was the best choice to make, he would have chosen C. That is, he had conditional alternatives open to him.

But this is not the only way in which to interpret this situation. The compatibilist can also state that the

agent had epistemic alternatives open to him. Although he was determined to choose B he did not know this at the time of his deliberation, thus he could have chosen A or C as far as he knew at the time. He had epistemic alternatives open to him because his future was unknown to him.

I know of only one unequivocal supporter of this account of alternatives, despite the fact that it seems to me to be the most obvious account for the compatibilist to adopt (as I discuss in chapter 6.) This is Dennett, who states:

The useful notion of "can," the notion that is relied upon not only in personal planning and deliberation, but also in science, is a concept of possibility - and with it, of course, interdefined concepts of impossibility and necessity - that are, contrary to first appearances, fundamentally "epistemic." (1984, p.148)

This concludes my outline of the conditions of libertarian and compatibilist free will. I stated in chapter 1 that there are three main issues that need to be tackled in the free will debate. Firstly, whether our experiences of freedom point us towards any particular conception of free will; secondly, whether we can fulfil the conditions necessary for libertarian free will, and thirdly, if we cannot, and if our experiences also point us towards compatibilism, whether a satisfactory compatibilist conception of free will can be expounded.

I have now tackled the first issue and concluded that our experiences can point us towards either one of two very different conceptions of free will: one that accords with libertarianism, the other with compatibilist. Having just outlined the conditions for libertarian free will and

compatibilist free will I am now in a position to discuss the second issue: whether we can fulfil the conditions necessary for libertarian free will.

## Chapter Three

### The U-condition

In this chapter I shall discuss whether we can fulfil the U-condition. In the first section of the chapter I shall consider the U-condition as it applies to free will and in the second section I shall consider it as it applies to moral responsibility, where somewhat different arguments apply.

#### 3.i The U-condition of Free Will

The U-condition states that to be free an agent must be the ultimate originator of his decisions. That is, the origin of his free decisions cannot be traced back through a deterministic chain of events. Ultimate origination therefore requires some indeterministic events to have occurred in the decision making process. According to the majority of libertarians (e.g. Thorp, 1980, p.2; Chisholm, 1982, p.24; van Inwagen, 1983; p.205, Trusted, 1984, p.57; Clarke, 1993, p.193) indeterminism alone is not sufficient to fulfil the U-condition, for then we would have to equate free will with randomness, which would be as unacceptable as equating it with determined freedom (one exception to this position is Robert Kane (1985) whose views I shall discuss later.) However, this position provides libertarians with a problem, as John Thorp has realised:

Having insisted that a necessary condition of freedom is natural indeterminism it seems unconsciously difficult for them to say what the sufficient conditions may be; freedom is, after all, something more than mere randomness; when libertarians try to say what this something more is, their talk becomes at best evasive and obscure, and at worst incoherent. (1980, p.2)

In expounding ultimate origination the libertarian is faced with the problem of explaining how we are able to originate our decisions in such a way that they are not determined by prior events, and yet are not random. He is, in effect, asking us to accept a further type of event on top of caused events and random events: what I shall term free\* events. (The asterisk is to distinguish such events from other, specifically compatibilist, senses of the term.) Trusted, for example, states:

..the essential feature of the libertarian's case is that ordinary human actions are not like ordinary physical events. (1984, p.95)

Although the problem is widely recognised, the solution - the mechanism by which free\* events can occur in a world apparently comprised of only caused and random events - has always proved difficult to determine. Consequently, most libertarians have offered little or no detailed explanation as to how we can originate our decisions (e.g. Chisholm, 1966; Taylor, 1966; Denyer, 1981; van Inwagen, 1983; Trusted, 1984; Clarke, 1993; and Lucas, 1993.)

Libertarians therefore rest their case very firmly upon a stout rejection of the other two positions: compatibilism and the position (more properly, positions) which rejects free will completely. They argue that because free will is incompatible with determinism, and

because the only other alternative - that we do not have free will - has too many dire consequences, that we must accept the libertarian thesis, even if the mechanism by which we originate our decisions remains mysterious.

I do not believe that this is an acceptable state of affairs. Even if the consequences of life without free will are objectionable, and even if compatibilism is unsatisfactory, these two factors on their own do not make libertarianism correct. They may make it more acceptable, but that is a different thing entirely. And if we do not find the idea of life without free will objectionable (e.g. Honderich, 1988b; Breer, 1989), or if we do not find compatibilism unsatisfactory, then the fact that the mechanism by which it is claimed we can originate our decisions remains utterly mysterious can only serve to increase our doubts about its adequacy.

But is it fair to expect the libertarian to be able to explain origination? Richard Taylor, for example, has claimed that the libertarian should not be expected to be able to explain how the agent can originate his acts since this would presuppose determinism:

To give an analysis of agency or the sense in which an agent is the cause of his actions would amount to giving an analysis of an act, an analysis which would of necessity presuppose the truth of a metaphysical presupposition that is not only dubious but probably false [i.e. determinism]. (1966, p.112)

Consequently:

That anyone who happens to defend a thesis of indeterminism should be unable to give an informative analysis of an act is therefore no difficulty or embarrassment for him. It is only a logical consequence of his position. (ibid, p.95)

It is perhaps pertinent to note that few libertarians seem to share Taylor's view. Pace Taylor, I fail to see why the nature of our actions should be inherently mysterious, such that any analysis is impossible, or why such an analysis should presuppose determinism. The mere fact that our actions take place, that they happen, must surely mean that there is some way or means by which they happen, they do not 'just happen' in some random way, and it surely follows from this that it is logically possible to analyse them, even if, in practice, it is quite beyond our abilities to do so at the present time - and all this without the need to presuppose determinism. Therefore I think that it really is in the libertarian's best interests to provide as full an account of origination as he can. I now wish to consider how he might go about this.

The basic libertarian position (with the exception of Kane) is that in order to originate our decisions we need free\* events, but neither determined or random events, or any combination of these, appear to be sufficient for the occurrence of free\* events. This raises two related problems for the libertarian. The first is a conceptual problem: what exactly are free\* events? How much sense can be made of this notion? The second is an empirical problem: regardless of how much sense can be made of free\* events, how can they possibly occur in the physical universe? To be more specific: how can they arise from the purely determined and random neurological processes of the brain? And if they do not, if they arise instead in some non-physical mind, as some libertarians have claimed (Swinburne, 1986), then there are a whole host of other

problems related to mind-body dualism that must be considered.

The first problem for the libertarian, then, is to specify the nature of a free\* event, and this is no easy task. For just how much sense can we make of the idea of an event that is neither determined nor random? It seems almost a truism that something is either caused or uncaused, that is, random. However, as Lewis Wolpert (1992) has recently argued, time after time common sense has been shown by science to be misleading or incorrect, and so it could be the case that this bit of common sense - that something is either caused or random - is also mistaken. It could be the case that there are free\* events, even if from the point of view of common sense this appears unlikely, and so we should be wary of being too dismissive of the libertarian's position. (However, at the same time it cannot be ignored that the current scientific view is that no such events exist, a point to which I shall return later.)

We could follow the lead of a number of libertarians and invoke the notion of agent causality in order to clarify the notion of free\* events (e.g. Chisholm, 1966; Taylor, 1966; Thorp, 1980; Clarke, 1993.) We could argue that free\* events are not actually uncaused. Whilst they do not exist as part of some deterministic chain of events, nor do they occur indeterministically like quantum events. They are in fact caused by the agent, or the self. We therefore need to allow that there are two fundamental types of causality, event causality and agent causality. However, whilst this removes the problem of how free\*



events can arise from caused and random events - they don't; free\* events are only possible through agent causality - it leaves the conceptual problem largely untouched. That is, the notion of agent causality provides no insight in itself into the nature of free\* events.

One libertarian who is conscious of this problem, John Thorp, has commendably tried to clarify the notion of agent causality. He states:

At moments when one churns over this very abstract subject it may seem that in being asked to believe in radical agent causality or self-movement one is being asked to believe in something which is strictly inconceivable like - at worst - an event which causes itself. (1980, p.102)

This is indeed the obvious conclusion to reach: if the agent's decision is not caused by a previous event, and yet if it is not random - if it is still in some sense caused - then what else is there but this curious notion of an event causing itself? Thorp tries to extricate himself from this situation in the following way:

..it would not be possible to allow the idea of a self-causing event, that is an event which is the same event as the event which caused it. Event causality is, if you like, nonreflexive. But that is not what is required here; what is required here is subtly but importantly different. We do not require that an event be the same event as its cause, but that an event be the same as its being caused. (ibid, p.102)

Thorp continues by saying that this 'sounds like the most grotesque logic-chopping of the Schoolmen' and he is right, although there is more to this than mere word play. What Thorp wishes to illustrate is how two events, the cause and the effect, can be one event, without stating that an event is causing itself. By describing an event as being the same event as its being caused he believes that

he is able to achieve this.

However, I do not see why this description should not apply to many or even to all cases of cause and effect, and how, therefore, it can help us to understand the nature of agent causality. Consider the situation in which one billiard ball strikes another and causes it to move. Here we have an event, the second billiard ball moving, with a clear cause, the first billiard ball striking it. But the description of the second billiard ball as being caused to move describes both events as one event. The description includes the cause of the movement by stating that the second ball is being caused to move by the first ball, but at the same time it also includes that movement, otherwise it could not be caused to move. We have two events describable as one event. And yet we would not wish to suggest that there was anything unusual taking place here. How, then, can this help us to understand the nature of agent causality?

There is a further problem. By invoking the notion of agent causality to explain free will the libertarian is asking us to accept the existence of a new form of causality, fundamentally different from the more established notion of event causality. But why should we accept the existence of this new form of causality, especially since it does not seem to be able to shed any light upon the nature of free\* events? Even Thorp seems to have his doubts about its adequacy:

The libertarian...is proposing a kind of causality which is not capable of being analysed into our normal kind of causality, event causality. It may seem that he is thereby wantonly introducing mystery into the world. I think that this charge is correct, but perhaps it can be rendered less damaging by the following anodyne

consideration. He is not so much introducing mystery into the world as introducing more mystery into the world: the event causality with which we seem so comfortable is itself unfathomably mysterious, as any glance at a freshman metaphysics text will show. He is not introducing mystery alongside clarity, but mystery alongside mystery. Still, one mystery is better than two, and it is one of the debit points for the libertarian scheme that it must propose additional primitives. The libertarian however thinks that the benefits here outweigh the costs. (ibid, p.106)

Unlike Thorp I do not think that the idea of event causality is nearly as mysterious as he claims, especially when one considers the sort of detailed analysis of event causality that it is possible to provide, detail which is almost entirely lacking in analyses of agent causality (e.g. Honderich, 1988a; Honderich actually defends the notion of nomic connections, but I take this to be equivalent to the event causality that Thorp describes.)

But even if event causality is ultimately as mysterious as agent causality, event causality has one clear advantage over agent causality: it is consistent with the current scientific world view. This is because event causality, unlike agent causality, does not violate one of the basic principles of modern science: ontological reductionism. This is the thesis that everything in the world is ultimately composed of nothing more than the entities of physics. All entities reduce to those of physics. And if everything is composed of nothing but the entities of physics, then everything must surely interact in a way that is consistent with the law-like, indeterministic way that the entities of physics interact. Given this, then agent causality is clearly unscientific, since agent causality, unlike event causality, is not

consistent with the law-like, indeterministic interactions of the entities of physics.

When faced with this problem there are three basic options open to the libertarian. The first is to simply accept its conclusion and argue that scientific unacceptability does not necessarily prove that the libertarian's position is wrong. It may be somewhat undesirable, in that, in the words of Roderick Chisholm, it implies that 'there can be no complete science of man', (1966, p.24) but this does not mean that agent causality is wrong, only that it is apparently beyond the domain of science. This position should not, perhaps, even be regarded as unscientific, for, after all, many scientists are perfectly happy to accept that God is beyond the domain of science. Why should the same state of affairs not apply to agent causality? On the other hand, such an option is hardly a great recommendation for agent causality, especially not for the libertarian who wishes to take libertarianism away from the incoherence and evasion of its past, and it does not measure up well against the second general condition of free will: empirical plausibility.

The second option open to the libertarian is to argue that science does not presuppose such a reductionist metaphysics (this option is favoured by Clarke, 1993, p.197-199.) Consequently, he can argue that mysterious though agent causality may be, it is not necessarily anti-scientific, nor, indeed, need it be a hindrance to a complete science of man.

However, pace Clarke, it seems quite clear to me that

science is ontologically reductionist, and this appears to be the view of the majority of scientists. For example, the physicist John Barrow states:

Ontological reductionism maintains that there is no 'elan vital'. All the material content of the world can ultimately be reduced to elementary particles and forces of the sort studied by physicists. Most scientists assume this to be true. (1988, p.304)

And if all the material content of the world can be reduced to the elementary particles and forces of physics then surely everything in the world must interact in a way that is consistent with the way that these particles and forces are taken to interact by physics. Nor is this reductionist belief confined to physicists. The zoologist Richard Dawkins states:

The heirarchical reductionist...explains a complex entity at any particular level in the hierarchy of organisation, in terms of entities only one level down the hierarchy; entities which, themselves, are likely to be complex enough to need further reducing to their own component parts; and so on. (1986, p.13)

Many philosophers, and some scientists, are wary of reductionism, and with certain sorts of reductionism they have good cause. But I wish to stress here that I am only concerned with ontological reductionism. This is the comparatively modest view that everything in the world is fundamentally composed of, and therefore interacts in the same manner as, the fundamental entities of physics, which at the moment basically equates with quarks and leptons (Davies, 1983, ch.11.) What I do not wish to defend are the sorts of reductionism which claim that all explanations can be cast in the language of mathematical physics, or that laws formulated in one area of science can always be reduced to special cases of laws in other

areas of science, or that mental states, such as beliefs and intentions, can be reduced to neurophysiological processes, or that all properties can be reduced to the properties of physical entities. Some of these reductions may be possible, but they are quite separate from the issue of ontological reductionism and are not the concern of this thesis.

Given that the majority of scientists, and many philosophers, are ontological reductionists the question that naturally arises is why? The answer is basically that all the scientific evidence points to this fact. The higher sciences do not seem to use any entities which are not aggregates of the entities found in the lower sciences, and, ultimately, physics. But the empirical evidence is not the only factor that contributes towards a belief in ontological reductionism, for the fact of the matter is that if the different individual sciences did not relate to each other in this way - if they contradicted each other over the ontological nature of the world - then their authority would be seriously undermined. In other words, science has to assume that ontological reductionism holds, and not just accept that it **may** be the case. The physicist Bernard d'Espagnat expresses these views when he states:

Nowadays, chemists would be among the first to acknowledge that it is in the field of physics, and more precisely of quantum physics, that the theoretical and conceptual foundations of their discipline are to be found. Similarly, most astronomers now see themselves as astrophysicists, and with good reason. It would be easy to make further observations of such a kind. In this sense the unification of the empirical sciences, long seen as an ultimate goal, is well on the way, under the leadership of physics. (1989, p.23)

But even if we believe that science should not presuppose that ontological reductionism holds, and should only be lead by the empirical evidence, then we still have to accept that there is simply no consistent empirical evidence for the claim that there are material entities in the world which are not composed of, or do not interact in the same manner as, the entities of physics. Those entities which do appear to defy such an analysis are either so lacking in evidence that their existence cannot seriously be maintained, or shown at a later date to reduce to physics after all. Examples of the former include various supernatural and paranormal entities - ghosts, telepathic forces, Cartesian minds, etc - one example of the latter is living tissue, which, until the rise of biochemistry this century, had always been considered to be fundamentally different in it's composition to non-living matter.

Some critics are apparently not convinced by such arguments, however. Tim Crane and D.H. Mellor have recently argued that reductionism is not possible because the physical sciences are not unified. They state:

The world even of the admittedly physical sciences contains a vast number of very different kinds of entities, properties and facts. That is why so many different sciences, using different methods, are needed to study them. No one could think astrophysics and genetics unified even in their methods, except under the most abstract descriptions of scientific methodology. And in their contents, they display no more unity than that of a conjunction. (1990, p.188)

It is indeed debatable whether the different sciences are unified in their methods, but this is quite irrelevant to the question of their ontological unity. As to their contents, it is difficult to see how Crane and Mellor's

argument can be justified. The world of the physical sciences does indeed contain a vast number of different entities, but there are no entities in it which cannot be seen to be composed of more simple entities of the sort found in physics. Astrophysics and genetics may be very different in their subject matter but they are both concerned with entities which appear, ultimately, to be the same.

Genetics is basically concerned with the study of DNA, RNA, and allied entities. DNA is a long chain made up of a various different molecules, which are themselves composed of atoms, which in turn are made up of protons, neutrons and electrons, and ultimately, according to current physics, quarks and leptons. Astrophysics, by contrast, is concerned with the study of stars, and other stellar material, which are mostly composed of hydrogen and helium gases at very high energies. But hydrogen and helium gases are also composed of protons, neutrons, and electrons, and therefore, ultimately, quarks and leptons. So astrophysics and genetics both deal with objects which are composed of the same fundamental entities. In their contents they display a good deal more unity than that of a conjunction.

Crane and Mellor are also critical of what they term 'microreduction', 'the idea that there is really no more to things than the smallest particles they are made up of,' (ibid, p.189). It is not clear in exactly what sense they mean this phrase to be taken, for whilst it can be taken to imply ontological reductionism, it can also be taken to imply a great deal more than that. If the phrase



is taken in its most radical sense, for example, then it could be taken to imply, amongst other things, that since there is really no more to our behaviour than the complex interaction of quarks and leptons, and that since these entities do not exhibit, say, rationality, that it must follow from this that we do not exhibit rationality either, but are as mechanical in our behaviour as the entities from which we are composed. It should be clear that the ontological reductionist would in no way wish to support such a radical microreduction, for clearly, much of the time, we are rational in our behaviour, and such a view is perfectly consistent with ontological reductionism.

If the phrase is taken to imply only ontological reductionism, then the examples which Crane and Mellor provide to refute their unspecified microreduction simply do not affect it. They state that some physics 'is positively macroreductive: Mach's principle, for example, which makes the inertial mass even of microparticles depend on how matter is distributed throughout the universe.' (ibid, p.190) But Mach's principle in no way implies that, for example, the mass of a body is not fixed by the mass of its individual constituents. It does not affect ontological reductionism.

Crane and Mellor's next example concerns a gas sample:

Suppose for instance that our sample's volume is suddenly halved at a constant temperature. If the gas is ideal, Boyle's law entails that when its pressure settles down again it will be twice what it was. That law does not dictate all the interim behaviour of the sample's molecules - except that it must be such as will eventually double the sample's pressure. That much of their behaviour

is determined - and thereby explained - macroreductively by a law governing the sample as a whole. (ibid)

This again does not affect ontological reductionism. It does not suggest that the behaviour of the gas is not fixed by the behaviour of it's molecules, or by the behaviour of the entities from which those molecules are composed. Boyle's law, and other 'macrolaws', may have greater authority than some of the microlaws used to explain the behaviour of microparticles, but this is simply because these microlaws do not (yet) have the experimental success of Boyle's law.

Crane and Mellor's final example concerns quantum theory. They state that although it is sometimes true that facts about parts explain facts about wholes, it is not always true, even in microphysics:

If for example we take the quantum mechanical description of a quantum ensemble to be complete (as orthodox interpretations do), the superposition principle entails that its properties will not be a function only of those of its isolated constituents plus relations between them. (ibid)

However, the problem with this argument is that it relies upon a theory - orthodox quantum theory - which is itself severely defective ontologically. Quite simply, orthodox quantum theory fails to provide a realistic answer to the problem of just what sort of entities electrons, protons, photons, and so on, are, in view of the contradictory wave and particle properties that these entities appear to possess:

QT [Quantum theory] is above all a theory which is about micro objects or systems, and which seeks to predict and explain macro phenomena in terms of micro phenomena. If QT is to achieve this, it is clearly important to develop QT as a theory which can be interpreted micro realistically, like classical theories, as being about micro systems evolving and interacting in space and time - a theory with its own definite, characteristic physical

ontology. But OQT [orthodox QT] cannot be given such a micro realistic interpretation, just because OQT provides no solution to the wave/particle problem. (Maxwell, 1988, p.2)

Many physicists deny that quantum theory needs to provide a solution to this problem, they deny that it needs to take a realist stance, and hold instead that physics should only be concerned with developing theories which merely predict more and more phenomena more and more accurately. By taking this more modest instrumentalist line, they avoid the charge that quantum theory is defective because it cannot tell us what sort of entities electrons and protons are. But quite apart from the fact that this line of response blatantly ducks what is a fundamental problem, and one to which we should endeavor to find an answer, there are a number of other basic problems with even an instrumentalist quantum theory (Maxwell, 1988, pp.3-6) which show that it would be quite inappropriate to conclude, from the fact that orthodox quantum theory conflicts with ontological reductionism, that we have good reason to reject ontological reductionism. It is orthodox quantum theory which should be questioned.

It seems to me that there are no good reasons for doubting that science is ontologically reductionist. In which case, given that science presupposes that everything is ultimately composed of, and interacts in a manner consistent with the interaction of, quarks and leptons, it must be concluded that agent causality, being, as the libertarian claims, a fundamentally different sort of causality to event causality, is not consistent with the

current scientific picture of the world.

There is a final option open to the libertarian when faced with the problem of the scientific acceptability of agent causality. This is to argue that even if science does presently presuppose a reductionist metaphysics, it should not. It should recognise the validity of agent causality. This is Thorp's position:

It is not really in its overt claim for causal determinism that the scientific world-view conflicts with human freedom (for these claims are no longer always upheld) but rather in its much more covert claim - its undeclared and unexamined presupposition - that the only causality is event causality. What the libertarian must do, it seems, is uncover and challenge this presupposition of science. (1980, p.104)

In other words, it may be that the current scientific view of the world is incomplete, and that it should recognise that agent causality is a valid form of causality. It could be that the neurons of the brain somehow enable agent causality to arise, even though they themselves are subject to event causality. How plausible is this argument? Could free decisions somehow arise from the processes of the brain, and yet be governed by agent causality and not event causality? This is what Thorp believes. He believes that it is possible to develop a coherent account of agent causality within a non-reductive identity theory. He states:

Richard Taylor, in his book 'Action and Purpose', proposes what he calls agent causality. According to this theory the immediate cause of a human act is the agent, not some event or other. He argues that the idea of an agent as cause is the original and more natural causal conception; the theory that only events can cause events is a new and less natural theory which there is no compelling reason to accept. (Thorp, 1980, p.99)

According to Thorp 'This proposal is an elegant and appealing one' (ibid, p.99). He states:

A decision is an event. We asked what caused it and the only possible answer seems to be either that another event caused it - in which case we are not free, or that nothing caused it - in which case we are not free but in the bondage of randomness. Events indeed are either random or else caused by other events; there is no sense at all to the idea of a self-caused event. On the other hand if the libertarian can loose himself from the grip of the idea that only events can cause events, then he can perhaps admit that sometimes objects can cause events. And then when he asks what caused a given decision he would no longer be caught between the equally embarrassing alternatives that another event did or that nothing did - it just happened. He can find a middle ground between determinism and randomness by saying that the decision was caused by an agent. (ibid, pp.99-100)

However, by basing his account around an identity theory Thorp is immediately faced by the problem of explaining how agent causality can arise from a brain whose neurons interact through event causality. For according to the identity theory, even a so-called non-reductive identity theory, mental events and brain events are in some sense the same. Surely, then, the mind state is always dependent upon the brain state, because the mind state is the brain state. Thus mind events, such as decisions, are always dependent upon brain events. In which case agent causality cannot help but be a case of event causality.

Thorp is conscious of such an objection and tries to counter it in two ways. Firstly he argues that the mind state is not always dependent upon the brain state, but that sometimes the brain state is dependent upon the mind state, even within an identity theory of the mind. Consequently agent causality need not be dependent upon

event causality. Secondly, he argues that due to quantum indeterminism within the brain, agent causality is not analysable in terms of event causality. Consequently agent causality should be regarded as just as fundamental a form of causality as event causality. I shall consider each of these points in turn.

Thorp's claim that the brain state is sometimes dependent upon the mind state relies upon introducing the notion of 'hegemony'.

Often...in a system of events bearing two levels of description the one description has what we might call explanatory priority over the other; that the event occurred under the explanatory prior description explains its occurrence under the other description, but not vice versa. 'Explanatory priority' is a cumbersome expression; I shall use in its stead the term 'hegemony'.

Hegemony, then, is a property which can be possessed by one of two (or more) descriptions of an event, such that the event's occurrence under the latter description(s) is explained by its occurrence under the former description. We could stipulate also that in a case where each description seems to explain the other(s), no hegemony is to be ascribed. That is, a description is hegemonic if (a) it explains and (b) it is not explained by, other descriptions of the same event. (ibid, p.86)

Thorp goes on to state that hegemony can be applied to the mind-brain relationship. However, one cannot fail to notice that as he continues his explanation he subtly changes the relationship between mind and brain that he is describing from a descriptive one to a causal one. This is because whereas hegemony is about the relationship between descriptions of events, what Thorp is really concerned with is the relationship between those events themselves. He wishes to 'offer a theory according to which it will seem reasonable to say that, sometimes, the brain state is dependent on the mind state.' (ibid, p.83) By introducing

the notion of hegemony, that it is sometimes possible for the mental to have explanatory priority over the physical descriptively, Thorp is able to prepare the ground for the more contentious, and quite different, view that the mental sometimes has priority over the physical causally. The following passage illustrates this change from description to causation.

It is plausible to suggest that when, for example, we walk in the country and allow our thoughts free reign, it is the neurology beneath them that, according to its own laws, directs their course. On the other hand when we force our thought onto some track, to perform a deduction, or to solve a problem, the stream of mental events is directed by some mental logical laws, and the mental descriptive level is hegemonic: here the mental descriptions drag the neural descriptions about according to the laws of sequence which belong to the mental: in the case of free associative thought the neural descriptions drag about the mental descriptions according to the laws of sequence (causal laws) which belong to neuralia. (ibid, p.91)

Consider his statement that the 'mental descriptions drag the neural descriptions'. This drag is the key word. It describes a causal relationship and not just an explanatory one. Thorp is not just saying that sometimes a mental description is more explanatory than a physical description, he is saying that sometimes the mental description 'directs' the physical description. That is, that the mental sometimes has control of the physical. The introduction to the chapter following his discussion of hegemony makes Thorp's position clear:

In the previous two chapters we have shown how the state of the brain may sometimes lack sufficient physical causal conditions, and we have shown that at those times the brain state may be dependent on the mind state and not vice versa. Further, we have been able to show this while adhering to a fairly conservative theory of mind, the Identity theory. (ibid, p.95)

Thorp believes that he has shown, via hegemony, that sometimes the brain state is dependent upon the mind state, and that this is possible within an identity theory of mind. But it seems clear to me that all he has actually shown with his notion of hegemony is that sometimes a brain state **description** of an event may be dependent on a mind state **description** of an event as regards explanatory priority. In other words, the best way to understand certain brain processes may be by reference to their corresponding mental processes - the agent's intentions, beliefs, desires, etc - and not purely by reference to other brain processes. Hegemony is, after all, only concerned with descriptions, as Thorp himself states. But there is nothing in any of his arguments to suggest that the brain state is sometimes dependent upon the mind state in any way other than a descriptive one. He certainly does not show that the brain depends upon the mind causally.

Consequently, it remains completely unclear to me in what meaningful sense one can assert that the brain can be dependent upon the mind in a way that implies that the mind state is not reducible to the brain state **if one is working within an identity theory**. Even if Thorp is operating within a non-reductive identity theory then it still follows that the 'mind state' that he describes is identical, in some sense, with certain processes within the brain. Therefore, to state that the brain state may be dependent upon, or lead by, the mind state is, in identity theory terminology, to state that the brain state may be dependent upon, or lead by, certain processes within the



brain. And this in no way implies that the mind state is not reducible to the brain state. Clearly Thorp wishes to say something more substantial than this, but what? He explicitly states that he does not wish to argue for a dualism of mind and brain and yet how can the 'brain state' be dependent upon the 'mind state' in any meaningful way if the two are not separate, that is, if some sort of dualism does not hold? I do not see how. I therefore conclude that Thorp's first argument to show that agent causality is not dependent upon event causality fails.

Thorp's second argument is based upon the claim that agent causality is not analysable in terms of event causality because of quantum indeterminism within the brain:

It is the events in the nervous system which would constitute the most detailed train of causally linked events leading up to an action. If this train of sufficient causes is broken just before the action (or decision), then the proposed analysis of the alleged agent causality into event causality does not succeed. The events in the agent leading up to the action do not amount to sufficient conditions for that action. (ibid, p.100-101)

I do not find this argument convincing. Pace Thorp, there appear to me to be plenty of cases which show that the events leading up to an action can amount to sufficient conditions for that action even if some of them are indeterministic. For example, it has recently been shown that only a few photons of light are sufficient to cause the light sensitive cells in the eye to fire, and one cannot wish for better examples of indeterministic events than photons of light. And the indeterministic

decay of radioactive material is sufficient to activate Geiger counters. I therefore simply do not see why it follows, from the inclusion of indeterministic events within a chain of events, that that chain of events is not sufficient to cause the resulting action. Consequently, I do not see why this should prevent one from analysing agent causality in terms of event causality. I conclude that Thorp's second argument also fails.

I have now rejected both of Thorp's counter-arguments to the claim that agent causality is simply a certain sort of event causality (at least, as far as the identity theory is concerned.) Of course, this does not prove that radical agent causality, and therefore origination, is not possible within an identity theory, but it does cast much doubt on it. But Thorp's failure should not be surprising. For what he has to show is that a mind which interacts through agent causality can arise from a brain whose neurons interact through event causality. He has to show that it is possible for the self to be in some sense identical to the brain, whilst at the same time interacting in a fundamentally different way to the brain. In other words, he has to show how two things can in one sense be fundamentally different, and yet in another sense be the same. This seems to me to be asking the impossible.

In fact, it seems to me that the notion of agent causality naturally leads to some form of mind-body dualism. For in stating that free\* events originate in the self, and are fundamentally different to both random and determined events, the libertarian implies that the self is rather like an island of agent causality in a sea of

event causality. He implies that the two are in some significant sense distinct from each other, a state of affairs which is surely more likely to occur in a universe in which there are two fundamentally different realms - the mental and the physical - rather than in a universe in which everything is physical.

A number of libertarians are visibly critical of dualism (Thorp, 1980, p.120; Trusted, 1984, p.36, p.111), and yet I cannot see how a non-dualist stance would be able to avoid the problem just raised. On the one hand such a libertarian admits that the self is somehow fundamentally different to everything else, in that it alone is subject to agent causality, whilst on the other hand he wishes to argue that the self arises from the brain, in that he explicitly rejects a dualist explanation. This is not a problem which faces the dualist libertarian, for he accepts that the mental and the physical are fundamentally different. Consequently, not only should one not be surprised if they interact in fundamentally different ways, one should expect it.

However, such a dualist libertarian position has little else to recommend it, and it is not difficult to appreciate why some libertarians are critical of it. For, as is well known, dualism suffers from a variety of problems, which, whilst they do not refute it, certainly make it a very unattractive position indeed.

For example, despite the best efforts of John Eccles (1986, 1989) it is still unclear, first, how the physical and the mental can interact, second, where the mental is situated, and third, how it can have evolved (for fuller

criticism of dualism see Churchland, 1984; Smith and Jones, 1986; and Lycan, 1987). Furthermore, the overwhelming evidence from the neurosciences is that consciousness is due only to neuronal activity, and not to some external Cartesian mind (Churchland, 1986; Honderich, 1988a.)

A further point against this position, as the libertarian John Thorp has noted, is that the dualist libertarian is not able to escape the conceptual problem - the problem of what free\* events actually are - for this problem would arise

..even in a world of disembodied minds from which all causal necessity was wished away. Such a disembodied mind would be the subject of many events, all of them uncaused. Some of these would perhaps be free decisions. Even for the disembodied mind the question arises, what is freedom more than mere randomness? - for surely it must be more. (1980, p.95-96)

And the dualist libertarian is not able to answer this question any more satisfactorily than his materialist counterpart. Consequently, the dualist libertarian position fares no better than the materialist libertarian position as regards conceptual and empirical adequacy.

Most of the problems associated with origination arise because the concept of free\* events is so mysterious. If the libertarian were to abandon this concept then perhaps he would be able to formulate a more satisfactory account of origination. Just such an approach has recently been taken by Robert Kane.

Kane's account of free will basically holds that free events consist of random events occurring in an otherwise determined stream of events. For example, suppose that I am undecided between living in A-land, B-land or C-land,

and that after some (determined) deliberation I have only succeeded in rejecting C-land. Just then the indeterministic thought pops into my head that B-land has pleasant beaches, and this causes me to choose B-land (Kane, 1985, p.104.) Although indeterministic, Kane argues that my choice was rational because I had already decided not to count factors in favour of C-land, and because I like beaches. And since it clearly originated within me, and since, if the indeterministic thought had occurred in a different manner, I physically could have chosen otherwise, my choice appears to satisfy all the libertarian's criteria for a free decision.

In fact, Kane proceeds to build up an account of free will that is more sophisticated than this, but such detail will be unnecessary here. All I wish to illustrate is that Kane explicitly rejects the standard libertarian notion of origination, that is, couched in terms of what I have called free\* events.

Kane's belief that indeterminism alone is what is required for free will, rather than free\* events, arises from his belief that in order to be free we only need to fulfil what he terms the Condition of Sole or Ultimate Dominion (CSUD) over our decisions:

An agent's power (or control) over a choice at time  $t$  satisfies the condition of sole or ultimate dominion if and only if (i) the agent's making the choice rather than doing otherwise (or vice versa, i.e. doing otherwise rather than making the choice) can be explained by saying that the agent rationally willed at  $t$  to do so, and (ii) no further explanation can be given for the agent's choosing rather than doing otherwise (or vice versa), or of the agent's rationally willing at  $t$  to do so, that is an explanation in terms of conditions whose existence cannot be explained by the agent's choosing or rationally willing something at  $t$ . (ibid, p.46)

This condition, he argues, can be achieved by the inclusion of indeterministic events in our otherwise determined deliberation. He adds

I think that D3 [Kane's code for CSUD] expresses the requirement Bramhall and other libertarians had in mind when they talked about free agents being the sole or ultimate causes of their choices, or having ultimate control over the springs of their actions. The idea that a free choice must be solely and ultimately one's own product, as G.H.Hardie and Michael Slote have said, is what explains the significance of the expression "one's own" in the longer expression "of one's own free will." (ibid, p.47)

However, it seems to me that Bramhall, and most other libertarians, would not find it acceptable. For although the condition may be satisfied by including indeterministic elements within otherwise deterministic thought process, as Kane argues (although see Double, 1991, pp.203-207, for an objection), it still places the ultimate control of the decision in the hands of the indeterministic laws of physics, rather than in the hands of the agent. The agent may control his deliberation, he may control the options between which he randomly chooses, but he does not control the actual choice. And this would surely be quite unsatisfactory to the majority of libertarians.

After all, the very reason that libertarians reject compatibilism is because they believe that the decisions of a person with compatibilist free will will not ultimately be under that person's control: they will be under the control of the laws of physics. And simply adding an indeterminist element into an otherwise deterministic account of free will does not enable one to overcome this problem. The decisions of a person with this

revised free will will still be at the mercy of the laws of physics as much as if they were determined, and for the standard libertarian this will be unacceptable. It will simply not enable him to be in ultimate control of his decisions. Therefore, pace Kane, it seems clear to me that the condition of sole or ultimate dominion does not express the requirement Bramhall and other libertarians have in mind when they talk about free agents being the sole or ultimate causes of their choices, or having ultimate control over the springs of their actions.

Kane recognises that his account of free will only allows partial control of one's decisions but states that 'we should consider it a veiled clue to our limited condition as free beings' (1985, p.97.) But if the limit in question holds that we are limited to only being in control of our determined decisions, and not of our undetermined ones, then this is surely tantamount to admitting to the compatibilist (especially the compatibilist who argues that we need to presuppose determinism in order to be free) that his position is the correct one. Consequently, Kane's revision of libertarianism will be considered as unacceptable by libertarians and unnecessary by compatibilists.

This concludes my discussion of the U-condition as it applies to free will. I have argued that the libertarian appears to be unable to provide satisfactory answers to either the conceptual or empirical problems that it raises. Free\* events remain unavoidably mysterious and quite inconsistent with the current scientific world view. Even when the notion of agent causality is introduced, and

examined in some detail, the conceptual problem seems no nearer to resolution, and the libertarian is still faced with the problem of either explaining how the physical world can incorporate such a curious notion of causality or of opting for a dualistic view of the world with all its inherent problems. I conclude that there is no sound evidence to suggest that we can fulfil the U-condition of free will. This does not prove that it cannot be fulfilled, of course, but only that it remains a profound mystery as to how it can. Even so, it hardly amounts to a ringing endorsement.

### 3.ii The U-condition of Moral Responsibility.

The libertarian believes that since a free agent can originate his decisions that he can be ultimately responsible for his actions in a way which is inconsistent with the agent being purely a product of his heredity and environment. Otherwise, they believe, his heredity and the environment would be ultimately responsible for his actions, and he would not be responsible for them at all. Among those who explicitly defend such a condition of responsibility are C.A.Campbell (1957, p.164) and Paul Gomberg (1978, p.208), although the condition is less explicitly stated (but still clearly held) by a number of other libertarians, as Martha Klein (1990) has recently shown.

As Klein states, there are two separate issues at stake here:

1. whether the U-condition is logically possible



2. whether the U-condition, whilst being logically possible, is empirically possible.

She argues that the U-condition is logically possible but (probably) not empirically possible. In this section I shall argue against Klein that the U-condition is not logically possible (from this it follows, of course, that it is not empirically possible either.) My discussion will centre around an old and familiar argument against the U-condition, variations of which are found in Edwards (1961, p.121) and Gomberg (1978, pp.207-209.) I shall consider the argument in it's most recent formulation, as it is expounded by Galen Strawson (Strawson actually refers to true responsibility rather than ultimate responsibility but it should be clear that the two terms are equivalent.) Here is the argument:

(1) Interested in free action, we are particularly interested in rational actions (i.e. actions performed for reasons as opposed to reflex actions or mindlessly habitual actions), and wish to show that such actions can be free.

(2) How one acts when one acts rationally (i.e. for a reason) is, necessarily, a function of, or determined by, how one is, mentally speaking.

(3) If, therefore, one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking - in certain respects, at least.

(4) But to be truly responsible for how one is, mentally speaking, in certain respects, one must have chosen to be the way one is mentally speaking, in certain respects. (It is not merely that one must have caused oneself to be the way one is, mentally speaking; that is not sufficient for true responsibility. One must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, at least, and one must have succeeded in bringing it about that one is that way.)

(5) But one cannot really be said to choose, in a conscious, reasoned fashion, to be the way one is, mentally speaking, in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, 'P<sub>1</sub>' - with preferences, values, pro-attitudes, ideals, whatever - in the light of which one chooses how to be.

(6) But then to be truly responsible on account of having chosen to be the way one is, mentally speaking, in

certain respects, one must be truly responsible for one's having **these** principles of choice  $P_1$ .

(7) But for this to be so one must have chosen them, in a reasoned, conscious fashion.

(8) But for this, i.e. (7), to be so one must have already have had some principles of choice,  $P_2$ , in the light of which one chose  $P_1$ . (1986, pp.28-29)

And so on. Strawson concludes that ultimate responsibility 'is logically impossible because it requires the actual completion of an infinite regress of choices of principles of choice.' (ibid, p.29)

According to Strawson the argument is likely to be resisted at two points, (3) and (4). He states:

It may be objected that one does not have to be at least partly truly responsible for how one is, mentally speaking, but only for how one decides; and that one can make a fully deliberate decision and be truly responsible for it even if one's character, say, is entirely determined (or entirely not self-determined). (ibid)

But such an objection seems to me to be a non-starter. How can the libertarian possibly hold someone ultimately responsible for his (undetermined) acts if they were to arise from his determined character? He would have to assert that there is some fundamental distinction between a person's character and their decisions such that whilst the person is not ultimately responsible for their character (because it is determined) they are ultimately responsible for the decisions which arise from that character. I fail to see how such a situation could arise. If our decisions do not come from our character then where do they come from? A person who consistently makes morally objectionable decisions is held to do so because he has a morally objectionable character. Although we may accept that certain people of good character do occasionally make morally objectionable decisions, we would surely not wish

to regard someone who consistently did so as still having a good character? And if decisions do come from our characters, and we are ultimately responsible for our decisions, then surely it must be because we are ultimately responsible for our characters as well? In which case they cannot be determined (according to the libertarian).

Strawson states that (4) may also be objected to: 'in particular the idea that one must have chosen to be the way one is.' (ibid) He says that it seems an absurdly artificial condition to place upon true responsibility, but, he continues 'one should ask oneself what else being at least partly truly responsible for how one is, mentally speaking, could possibly consist in.' (ibid)

This condition is not as absurd as Strawson thinks. Consider the following quote by Karl Popper:

To a certain degree the personality somehow really does form itself actively. Admittedly, it may be partly performed by its genetics. But I think that we both believe that this is not the whole story, and that a great part of the formation is really achieved by the free actions of the person himself. (Popper and Eccles, 1977, p.472)

Although it is possible for a compatibilist to make this claim he would still recognise that the personality formed by the person's free actions was itself only a product of its genetic inheritance and the affects of the environment upon that inheritance (in the form of the brain). But I think that it is quite clear that Popper, a staunch libertarian, means much more than this. For Popper the (non-physical) self really does form itself in a way that is above the formation of a self by genetics and environment alone and which is therefore consistent with

it being ultimately responsible. It forms itself - and so it, and not just it's genetics or the environment, is responsible for its formation.

Furthermore, as Strawson says, what else could true or ultimate responsibility possibly consist in? Although they may not be as explicit as Popper it seems clear to me that most libertarians hold views echoing those given in point (4).

It seems likely, though, that Strawson's criticism of the U-condition will not be universally accepted. Paul Gomberg, for example, has raised two possible, if highly implausible, ways in which the sort of argument given by Strawson might be resisted, although he does recognise that the metaphysical position required by his proposed solutions will be found by most to be uncomfortable.

He states that there are two ways in which we could overcome the infinite regress problem in an argument like Strawson's:

Some philosophers have suggested that there might be an infinite causal chain that converges to a single point in time. If such a chain existed then every event in the causal sequence could be such that the agent is responsible for it. However, I know of no evidence that such causal sequences exist. (1978, p.209)

This argument seems to imply that the libertarian need not be bothered by an infinite regress argument, since we can conceive of an infinite causal chain disappearing into the past. But even if we granted that such chains exist - a big if - then I do not see how this could help the libertarian. For surely he needs to explain how responsibility first arose? And by positing an infinite chain he is saying that it never arose, but that

it just is, that it has always existed. It remains unclear how we should make sense of this.

Gomberg's second suggestion involves reincarnation:

For example, the law of Karma in Indian philosophy hypothesises that the soul has always existed and that what one is and does at any point in time is something that one is responsible for in virtue of one's actions and character in a previous life. (ibid)

But this still does nothing more than suggest that there may exist an infinite past. It does not explain why this should enable us to be ultimately responsible. The metaphysical position expounded here is not just uncomfortable, it is quite baffling.

A more recent objection specifically aimed at Strawson's argument has been provided by Ted Honderich (1988a.) The basic argument, he says, is that since a certain conception of responsibility requires that I stand in a certain creative relation to an action, and since that action has a certain source or ground, I must also stand in the given creative relation to the source or ground, and so on back through an infinite regress. But, says Honderich, libertarians need not accept that their conception of responsibility faces an infinite regress:

It can be asserted that the responsibility in question requires that I stand in a certain creative relation to an action, and hence in a certain creative relation to what it certainly has, a certain ground or source, but that either this ground or source or an earlier one is **primitive**. That is, it has no ground or source. An indeterminist theory of the mind may centre on this primitive episode. There will certainly be great difficulties about its description...but there need be no capitulation to a fiat, the fiat that any such episode must have a similar antecedent. (1988a, pp.179-180)

It is unclear just what Honderich has in mind when he talks of a primitive ground or source. But even if sense can be made of this (which Honderich himself seems to

doubt) I fail to see how he has resolved the problem. If my ultimate responsibility is grounded in a certain source then either I am ultimately responsible for that source or I am not. If I am not ultimately responsible for it then it is unclear why I should be held ultimately responsible for any actions that find their grounds in that source, and if I am ultimately responsible for that source, then it surely cannot be primitive. For if it is primitive then surely nothing is ultimately responsible for it, and so no ascriptions of ultimate responsibility can be applied to any actions arising from it. Consequently, I simply do not understand how Honderich's argument refutes Strawson's position (a conclusion which echoes Strawson's own recently stated position, although he disappointingly fails to provide any details of his objection to Honderich's argument (Strawson, 1989, p.11.)

It is therefore my conclusion that the U-condition as it applies to moral responsibility is not logically possible.

## Chapter 4

### The C-Condition

In the first section of this chapter I shall argue that the C-condition is superfluous to ascriptions of free will. Therefore it does not matter whether we can fulfil it or not. In the second section I shall consider a number of objections that have been raised against this argument. In the third section I shall argue that even if the argument of the first section can somehow be countered, that the C-condition is nevertheless clearly inconsistent with the nature of time as described by Einstein's General Theory of Relativity. Either we reject the C-condition or we reject Relativity.

#### 4.i A Problem for the C-condition

The C-condition, as it applies to free will, states that an agent's decision was free only if he could have decided otherwise, there and then, in the exact circumstances that he found himself. In this chapter I shall argue, following a famous argument of Harry Frankfurt's, that the C-condition is not a necessary condition of free will. Consequently, it is irrelevant to free will as to whether or not we can fulfil it: we simply do not need to. However, that said, my actual discussion will largely be concerned with the C-condition as it applies to moral

responsibility, since it is with reference to moral responsibility that the majority of philosophers have discussed it. Although I think that it would be more obvious to begin by discussing the C-condition with reference to free will this would only serve to complicate matters when I came to consider the various criticisms of Frankfurt's argument and had to translate arguments in terms of moral responsibility into arguments in terms of free will. In any case, the conclusions I shall draw concerning the C-condition and moral responsibility apply as much to free will.

Frankfurt's original argument stated that what he termed the Principle of Alternate Possibilities (henceforth PAP, which is equivalent to the C-condition.) is false. PAP states:

A person is morally responsible for what he has done only if he could have done otherwise. (1969, p.829)

He argued that we can conceive of counterexamples to this principle that show that there are situations in which people cannot do otherwise but in which we would still wish to hold them responsible. What Frankfurt failed to appreciate is that his argument also refutes PAP as it applies to free will. This is because his argument refutes PAP as it applies to decisions and not just as it applies to acts. Interestingly, all of those who have commented on Frankfurt's paper, whether in support or criticism, have also only applied his arguments to moral responsibility. In fact, some of them have ignored the fact that Frankfurt applied his argument to decisions and have only considered PAP as it applies to acts. Peter van Inwagen, for example, describes PAP in the following way:



Almost all philosophers agree that a necessary condition for holding an agent responsible for an act is believing that the agent **could have** refrained from performing that act. (van Inwagen, 1975, p.189)

The upshot of this is that the supposed refutation of PAP only refutes PAP as it applies to acts and not as it applies to decisions. Consider Dennett's version of Frankfurt's argument:

Jones hates Smith and decides, in full possession of his faculties, to murder him. Meanwhile Black, the nefarious neurosurgeon, who also wants Smith dead, has implanted something in Jones' brain so that **just in case Jones changes his mind** (and chickens out), Black, by pushing his special button, can put Jones back on his murderous track. In the event Black doesn't have to intervene; Jones does the deed all on his own. (1984, p.132)

This appears to refute PAP because although Jones could not have acted otherwise, because of Black, Jones went ahead and murdered Jones without Black's interference; it was as if Black was not there. If Black had not been present we would surely wish to hold Jones responsible for murdering Smith. Consequently, if he is there, but plays no part in Jones' deliberations, we should still hold Jones responsible, for everything took place as it would have done if Black had not been present.

However, there is an obvious response for the libertarian to make against this supposed refutation of PAP. He can state that whilst it refutes one version of PAP, it leaves a revised version untouched. In Dennett's argument Black only acts when Jones has decided not to murder Smith. Consequently the libertarian need only revise PAP to apply to decisions and not acts:

A person is morally responsible for an act only if he could have decided otherwise, even if he could not put

that decision into practice (i.e. even if he could not have acted otherwise.)

In this case Dennett's argument does not refute PAP since Jones could have decided otherwise, even if he could not have acted otherwise.

However, this revised version of PAP is refuted by Frankfurt's original argument. Here is the argument:

Suppose someone - Black let us say - wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something **other** than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way. (1969, p.835)

He continues:

Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it...Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it. (ibid, 836)

From this argument we can see that the ability to decide otherwise appears to be of no relevance whatsoever as to whether a person is morally responsible or not. Imagine two people on almost identical worlds making the same decision. One of the worlds has a figure like Black, called White, the other does not. In both worlds the events leading up to the decision are the same, in both worlds the people make the same decision, but in one of

the worlds, because of the presence of White, the person could not have decided otherwise. Is he therefore less morally responsible than the other person? It is not clear why he should be. White did not need to interfere and so everything occurred as it would have done if White had not been there, i.e. as it did occur on the other world. There seems to be no obvious answer as to why he should not be considered morally responsible that would not beg the question.

In order to be effective Frankfurt's argument requires that there is no difference between the situation in which Jones decides to A without Black being present, and the situation in which Jones decides to A whilst Black is present, but does not interfere (because Jones, unknowingly, does what Black wants him to do). Consequently, in order to refute Frankfurt's argument, it must be shown that there is some essential difference between these two situations. It must be shown that Black's presence does somehow significantly affect Jones' situation, even though he does not interfere, and even though, from Jones' point of view, it is as if Black did not exist. It is not obvious how this can be shown.

I grant that Frankfurt's example is somewhat artificial: many of those to whom I have expounded it have been apt to eye it very suspiciously. Apart from the fact that they really feel as if they can decide otherwise when they make decisions, and Frankfurt's example seems to make this feeling a sham, there is the feeling that his example does not appear to deal with 'normal circumstances' and is therefore not a universal refutation of PAP. It is still

the case, it is claimed, that under normal circumstances figures like Black do not arise, and so we can decide otherwise.

But such a response misses the point of the example. It does not just seek to show that if a figure like Black were around that we could not decide otherwise, but rather to show that it is logically possible that a person can be in a position in which they cannot decide otherwise, and yet in which we would still wish to regard them as morally responsible. It is logically possible that there are situations in which PAP is not a necessary condition of moral responsibility. Situations in which ascriptions of moral responsibility do not rest upon PAP, but purely upon other conditions. But, and this is the point, if we can describe the conditions of moral responsibility adequately without reference to PAP in this example, then we can surely do it in all situations, including real life situations? It surely follows from the example that PAP, or the C-condition, is completely superfluous? This is the problem that faces the defendant of PAP.

That said, there is one obvious criticism that can be made of Frankfurt's argument. It could be argued that whereas one can conceive of cases in which it is clear to Black that Jones is not going to **act** in the way that Black wants him to, because, say, Jones twitches when he decides that he is not going to act in the way that Black wants him to, and so lets Black know that he is about to act contrary to Black's wishes, it is not clear how Black could know that Jones was about to make a different **decision**. In other words, whilst it is clear that someone

could know whether a person was or was not going to perform a certain act because the behaviour accompanying that person's decision would give prior warning of their intentions, it is not clear what prior warning could be given that someone was or was not going to **decide** to perform a certain act. Put simply: how a person will act can be indicated by how they decide, but what indicates how they will decide? There seems to be no clear way for someone such as Black to know, until the person has decided, what they will decide. Even if Black is a mind reader, he will not necessarily know what Jones decision will be until Jones has made it, for there seem to be no prior warning signals for decisions, as there are with acts, to warn Black.

Whilst this argument is initially persuasive I do think that there is an answer to it. Imagine that Black has a device that can record the electrochemical patterns of Jones' brain processes (much like an EEG) and interpret these into thoughts. Black wishes Jones to decide to do A, and to do A, and so he monitors Jones brain to determine what decision he will make. When Jones comes to the point at which he makes a decision we can imagine a variety of possibilities. He may be convinced he must do A, and will not let anything convince him otherwise, or he may be quite sure he should do A, but may consider his other options, or he may be completely unsure as to whether he should do A or not. In the first case Black need not be too concerned about Jones's decision, for Jones is adamant that he will do A. In the second case Black needs to keep a tighter watch. But it may be the case that after

considering what to do he will still decide to do A. It is the third case which gives Black most concern. So he monitors all of Jones' thoughts. He monitors his deliberations; his reasons for doing A, his reasons for not doing A. If Black comes to believe that Jones' reasons for doing A will outweigh his reasons for not doing A then he will intervene, before Jones has a chance to decide not to do A, and make Jones decide to do A (we may imagine that Black's device can implant thoughts into Jones head as well as read them). However, as it happens, having considered his options, Jones decides to do A and does A. The point is, he could not have decided to not do A, for Black, monitoring his thoughts, would have intervened if he thought that Jones' deliberations were leading him to decide not to do A.

In this example, then, it is Jones' deliberations that give Black the prior knowledge that he needs in order to know whether Jones will decide to do A or decide not to do A. We may imagine that Black is quite wary, and so Jones only needs to show a few seconds of uncertainty for Black to intervene. Or we may imagine that Black is so familiar with Jones' style of deliberation that he has a very good idea (although not complete certainty) from the way in which Jones is deliberating of which way his decision will go. Either way, we may suppose that Black will be able to preempt Jones' decision before Jones makes it. He will therefore know when to intervene and when to sit back and watch. Consequently, he can bring it about that Jones could not have decided otherwise, in a certain situation, even though he need not intervene.

We could argue that Black only lets Jones make his own decisions if he is absolutely certain that Jones will make the decisions he wants him to. Otherwise he will intervene and make Jones decide what he wants him to decide. Consequently, on one of the rare occasions in which Jones decides to do something without Black's involvement (because he is certain that Jones will make the decision Black wants him to) PAP would be refuted. For, as I argued above, all that we need to prove is that there is at least one possible occasion when PAP is not a necessary condition of moral responsibility in order to show that we never need it. All that we need to show is that it is logically possible that there are occasions when the person is morally responsible, and yet PAP does not apply, in order to show that the condition is superfluous. In order to refute the PAP we do not need to show that Black will always be able to predict Jones' decisions, only that there is at least one occasion when he can predict his decision, and this does not seem to me to be an unlikely a state of affairs.

The only way in which someone like Black could be thwarted would be if it was impossible to determine from Jones deliberations how he will decide. But unless we accept an account of free will like Kane's, in which decisions arise randomly from determined deliberations, there seems to be no good reason to assume that Black would never be able to determine from Jones' deliberations how he will decide. Kane himself uses this argument to refute Frankfurt (1985, p.51.) However, in exchange for refuting Frankfurt's argument, we would then have to

accept Kane's deeply unsatisfactory account of free will. I presume that this is not an option that most libertarians would wish to take. Consequently, we can argue that Black would be able to prevent Jones from deciding otherwise if he thought that Jones might not make the decision that he wanted Jones to make, even if, in the actual case, he does not need to interfere, because it is clear to him that Jones will make the decision that Black wants him to make. And this is sufficient to refute PAP.

Following this argument the question naturally arises: could we adapt PAP again to overcome this problem? That is, in the same way that we can move from acts to decisions, to overcome the problems with PAP as it applies to acts, could we move from decisions to deliberations, to overcome the problems with PAP as it applies to decisions? Could we adapt:

A person is morally responsible for an act only if he could have decided otherwise, even if he could not put that decision into practice (i.e. even if he could not have acted otherwise.)

to apply to deliberations? Black would then be thwarted because he could not know how Jones' deliberations will proceed, in the way that he can know of his decisions. For whereas Jones' deliberations indicate what decision he will make there seems to be nothing that will indicate how his deliberations will proceed. The following somewhat clumsy condition suggests itself:

A person is morally responsible for the act which follows his decision which follows his deliberations only if he



could have deliberated otherwise, (even if he could not have decided otherwise.)

A moments reflection should highlight the shortcomings with this condition. Let us apply it to a real life situation: say, my going shopping. According to the condition, I am only morally responsible for the act which follows my decision to go shopping, which follows from my deliberation about whether or not to go shopping, if I could have deliberated otherwise. In other words, I am only morally responsible for the act which follows my decision to go shopping if I could have deliberated about something else. But this is clearly unsatisfactory: the something else could have nothing to do with shopping at all. In which case we are left with the condition that I am only morally responsible for the act which follows my decision to go shopping if I could have deliberated about, say, what shirt to wear. But this would make no sense. It would mean that someone who deliberated about whether or not to murder someone would not be held responsible for his act if he could not have deliberated about something else. But it does not follow from the fact that someone could not have avoided deliberating about whether or not to murder someone that they could not have avoided murdering them; it would not exonerate them of blame. For plenty of people deliberate about whether or not to murder someone and then do not. Therefore the libertarian cannot claim that just because someone cannot help but deliberate about whether or not to murder someone that they should not be held responsible for their action if they do decide

to murder the person.

The condition fails because adapting it from decisions to deliberations faces a fundamental problem. Whereas when I make a decision, or act, there is an alternative decision or act that I could have made which would still be relevant to the situation, when I deliberate about whether or not to do something, there is not an alternative set of deliberations which are relevant to the situation. For the deliberation, unlike the decision or the act, includes both my options - whether to do A or not to do A - whereas my decision to do A only includes that one option, it does not include my decision not to do A which can therefore be given as an alternative. But there are no viable alternatives to my deliberating about whether or not to do something. Consequently, PAP cannot be applied to deliberations.

Although the objection to Frankfurt's argument that I have just been considering strikes me as the most obvious one to make it is by no means the only one. Indeed, since Frankfurt first expounded his argument against PAP there have been a variety of objections raised against it, but, as I shall now discuss, I do not find any of them convincing.

#### 4.ii Objections to Frankfurt's Argument

##### 4.ii.a David Blumenfeld's Objections

Frankfurt, having discussed and dismissed PAP, suggested at the end of his paper that it should be replaced 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. (1969, p.838)

In other words, it is the person's intentions, wants, desires, etc, that are the relevant factors in responsibility ascriptions, and not whether they could have done otherwise. If the person wanted to do what they did, if they did it knowingly and willingly, then they can be held responsible, regardless of whether they could have done otherwise or not.

David Blumenfeld, however, has argued that PAP need not be replaced by Frankfurt's principle, but only by a revised version of PAP, which he terms PAP':

A man is not morally responsible for what he has done if he did it **because** he could not have done otherwise. This implies that the fact that a man could not have done otherwise exempts him from moral responsibility only if the factors that render it impossible for him to do otherwise actually bring his action about. This covers the counterexample and squares, I think, with our sense of the intention of the original principle; it merely takes care of a loose end. Furthermore, this amendment would not materially affect the arguments of philosophers who have relied on PAP to show that determinism and moral responsibility are incompatible. (1971, pp.341-342)

Frankfurt considered and dismissed this alternative principle in his own article (1969, pp.837-838) but Blumenfeld argues that Frankfurt does not provide sufficient reasons to abandon PAP' in favour of his own principle. He concludes:

..even if hard and soft determinists alike were to accept Frankfurt's counterexamples to PAP, there is no reason to suppose that either would find his alternative principle intuitively acceptable. In fact, I think it is clear that they would find PAP' acceptable and Frankfurt's alternative principle unacceptable. (1971, pp.343-344)

I agree with Blumenfeld that Frankfurt's rejection of PAP' in favour of his own principle is unsatisfactory. Despite this, though, I believe that PAP' is still unacceptable since one can formulate counterexamples to it in the same way that one can for PAP. The type of counterexamples I have in mind were first expounded by Eugene Schlossberger (1986) to refute PAP. However, since they differ in an important respect from Frankfurt's own counterexamples they are also applicable to PAP'. One counterexample is as follows.

Harry intends to slap his landlord and begins to do so. However, before contact is made, William presses a button on his machine which block's Harry's neural impulses and which causes Harry's arm to slap the landlord. Schlossberger states that 'it seems evident beyond dispute that Harry is no less culpable and no less morally answerable for the slap than he would have been had William not pressed the button.' (p.43)

But if Harry is culpable then this is a counterexample to PAP' as well as PAP, since Harry slapped his landlord **because** he could not have done otherwise. Schlossberger's example is a rather fanciful one but one can easily imagine a more realistic variation. Harry goes to slap his landlord and trips, flying into his landlord and striking him. Harry could not have done otherwise, for once he had tripped the Earth's gravity and the position

of his arm meant that he had to fall into his landlord and strike him. However, since Harry intended to strike his landlord it seems that he should be held culpable even though he did it because he could not have done otherwise. Note that he did not do it **only because** he could not have done otherwise. That is, he did not fall and strike his landlord without having the intention of striking him. Consequently, Frankfurt's own principle is not touched by this counterexample, whereas Blumenfeld's is, and so I suggest that both PAP and PAP' should be rejected as valid conditions for moral responsibility.

#### 4.ii.b Peter van Inwagen's Objections

Peter van Inwagen accepts that Frankfurt has made out 'a good case' for the falsity of PAP. However, he also believes that there are variations on PAP which can be formulated that, whilst not exactly like PAP, are still near enough to PAP to show that the ability to do otherwise is relevant to ascriptions of moral responsibility. Whereas PAP may not be immune to what van Inwagen terms Frankfurt counterexamples (such as the example of Dr. Smith and Mr. Jones,) van Inwagen argues that his variations are immune to such counterexamples.

Although van Inwagen discusses Frankfurt's argument against PAP as if it refers only to acts two of his variations on PAP are actually applicable to decisions as well as acts, and are therefore of relevance to the current discussion.

The two variations in question are themselves two different versions of what van Inwagen terms the principle of possible prevention, or PPP. He states that since there is disagreement as to whether the events for which we hold people responsible are particulars or universals, he has formulated two variations of PPP, one based upon the idea that events are particulars (PPP1), the other upon the notion that events are universals (PPP2). I shall consider each in turn.

PPP1 is defined by van Inwagen as follows:

A person is morally responsible for a certain event (particular) only if he could have prevented it. (1983, p.157)

To which he offers the following counterexample:

Gunnar shoots and kills Ridley (intentionally), thereby bringing about Ridley's death, a certain event. But there is some factor, F, which (i) played no causal role in Ridley's death, and (ii) would have caused Ridley's death if Gunnar had not shot him (or, since factor F might have caused Ridley's death by causing Gunnar to shoot him, perhaps we should say, "if Gunnar had decided not to shoot him"), and (iii) is such that Gunnar could not have prevented it from causing Ridley's death except by killing (or by deciding to kill) Ridley himself. So it would seem that Gunnar is responsible for Ridley's death, though he could not have prevented Ridley's death. (ibid, p.161)

Van Inwagen rejects this counterexample because it is inconsistent. It is inconsistent, he says, because the event-particular 'Ridley's death' that would have occurred if Gunnar had shot Ridley intentionally, is not the same event-particular as the event-particular 'Ridley's death' that would have occurred if Gunnar had been caused to shoot Ridley by factor F. (Van Inwagen regards two specific event-particulars x and y to be the same event 'if and only if x and y have the same causes.' p.160) But the event-particular 'Ridley's death' brought about only

by Gunnar (Ridley's death (G)) and the event-particular 'Ridley's death' brought about by Gunnar and factor F (Ridley's death, (F)) clearly have different causes, for one involves factor F. Consequently they are different event-particulars. Therefore the counterexample, which refers to the two event-particulars, Ridley's death (G), and Ridley's death (F), as though they were exactly the same event-particular, is inconsistent. Van Inwagen concludes that he cannot see how a counterexample can be formulated that can avoid such an inconsistency, and that PPP1 is therefore an acceptable variation of PAP.

The fault with this principle is that it relies upon the validity of van Inwagen's highly questionable method of differentiation between event-particulars. If one accepts his method of differentiation then one must accept that the counterexample is inconsistent, and that PPP1 is not refuted by it. If, however, one does not accept his method of differentiation, then the counterexample remains consistent, and PPP1 is refuted by it. So should one accept or reject his method of differentiation? Unfortunately, this is a problem which is not easily settled. My own personal feeling is that van Inwagen's method of differentiation is unacceptable because it is overly picky. There are, I suggest, certain event-particulars which are of sufficient similarity for us to want to say that they are the same. But if one accepts van Inwagen's method of differentiation then one also has to accept that no two event-particulars are ever the same. Consequently, whether one finds van Inwagen's method of differentiation acceptable or not depends upon whether one

thinks, or feels, that certain event-particulars can be the same, or that no two event-particulars are ever the same.

Why does van Inwagen's method of differentiation have the consequence that one will never be able to describe two event-particulars as the same? Basically because it does not restrict itself in its definition of event-particulars to the immediate cause (or causes) of an event, but rather it also allows, or even requires, one to bear in mind the various causes prior to the immediate cause, without setting any limit as to how far back one can go. The immediate cause of Ridley's death (G) and the immediate cause of Ridley's death (F) were one and the same - Gunnar's intentional shooting of Ridley. However, the cause of Gunnar's intentional shooting of Ridley in one case (Ridley's death (G)) was Gunnar's own wants and desires, whereas the cause of Gunnar's intentional shooting of Ridley in the other case (Ridley's death (F)) was factor F. It is by reference to these different causes that van Inwagen differentiates between the two event-particulars, even though they both had exactly the same immediate cause.

But if one is able to use non-immediate causes to differentiate event-particulars then one will always be able to trace back the two chains of causal events from two seemingly identical event-particulars to some past time where their respective causes differed in some way. Even if one takes the most extreme example - two identical worlds, in which everything that happens on one world corresponds perfectly with everything that happens on the



other - then one can still trace back the respective causes of two seemingly identical event-particulars from the two worlds to the actual point of creation of the two worlds, the causes of which, since such worlds would have to inhabit different areas of space, would have to be different. And if the causes of the ways in which the two worlds were created differed in some way, then no matter how similar the subsequent event-particulars of the two worlds, their event-particulars could not, according to van Inwagen's method, be described as the same.

Personally, I find such a conclusion ridiculous. For it suggests that event-particulars that appear to be identical in every respect, that appear to have identical causes and effects, will not in fact be so. As I said earlier, I regard this to be unnecessarily pernickety. A better method of event-particular differentiation would surely be just to differentiate event-particulars by reference to their immediate causes. In which case there is no difference between the event 'Ridley's death (G)' and the event 'Ridley's death (F)' and van Inwagen's counterexample is not inconsistent. Consequently, PPP1 is refuted by it.

I realise that this is by no means a decisive refutation of PPP1, as it depends upon the rejection of van Inwagen's method of event-particular differentiation, a matter which perhaps relies too much upon intuition for complete satisfaction. But if too much intuition is considered a weakness then remember that this is a weakness that my position shares with van Inwagen's. His defence of his method of event differentiation is also

heavily dependent upon intuition, as he freely admits (1983, p.160).

However, even if we accept that PPP1 is immune to counterexamples there are various criticisms that can be made of it. Fischer (1986), for example, argues that it does not provide the person with the sort of alternate possibilities which are implied by PAP. He provides a similar counterexample to Frankfurt to refute PAP. A man, Green, has had a device implanted in his brain by some scientists which enables them to control Green, if they should wish to, but which otherwise has no affect over him. Imagine that one day he sees a drowning child and after thinking about it decides to save the child. Now, as in the earlier case of Jones, Green could not have done otherwise, for if he had decided not to save the child the scientists would have twiddled a few knobs and made sure that the device in Green's brain would have made him save the child. But, like Jones, Green does appear to be morally responsible for saving the child, since the scientists did not intervene and the situation was as it would have been if the scientists had not existed.

With this example in mind Fischer states that rather than formulate PPP1 as follows:

A person is morally responsible for a certain event particular only if he could have prevented it

the alternate possibilities theorist would insist on something like the following interpretation:

A person is morally responsible for a certain event particular only if he could have brought about an event particular of a different type (as a result of an intention to do so). (1986, p.54)

Fischer continues:

Now, the Frankfurt-style counterexamples do seem to show this kind of principle false. Consider Green again...the alternate-possibilities theorist believes that, if Green is morally responsible for saving the child, then he must have had more than one genuine opportunity; that is, there must be more than one genuine, open possibility in which he acts freely. But note that in the example there is only one path along which Green acts freely, and it is just this sort of situation that the alternate-possibilities theorist finds antithetical to moral responsibility. (ibid)

Consequently, PPP1 is really of no use to the libertarian who wishes to defend the C-condition as a necessary condition of moral responsibility.

Van Inwagen's second variation, PPP2, is similar to PPP1 but assumes that events of the sort for which we want to hold people responsible are universals and not particulars. He defines PPP2 as follows:

A person is morally responsible for a certain state of affairs only if (that state of affairs obtains and) he could have prevented it from obtaining. (1983, p.171)

As before, van Inwagen formulates his own counterexample to this principle.

Gunnar shoots Ridley (intentionally), an action sufficient for the obtaining of Ridley's being dead, a certain state of affairs. But there is some factor, F, which (i) played no causal role in Ridley's death, and (ii) would have caused Ridley's death if Gunnar had not shot him (or had decided to shoot him), and (iii) is such that Gunnar could not have prevented it from causing Ridley's death except by killing (or by deciding to kill) Ridley himself. So it would seem that Gunnar is responsible for Ridley's being dead though he could not have prevented this state of affairs from obtaining. (ibid, p.173)

Van Inwagen rejects this counterexample. He states that we can take the words 'Ridley's being dead' that occur in the counterexample to denote the universal, U(Ridley dies). But, he argues, whilst it is indeed true that Gunnar could not have prevented U(Ridley dies) from obtaining, it is not true that Gunnar is responsible for U(Ridley dies). It is obvious why people might think he

is. He did something (shot Ridley) that was sufficient for U(Ridley dies), and he did this intentionally, knowing full well that it was sufficient for U(Ridley dies). However, according to van Inwagen this argument is invalid. His reasoning is as follows: consider the state of affairs U(Ridley is mortal). When Gunnar shot Ridley, he performed an act sufficient for the obtaining of this state of affairs. But it would be absurd to say that Gunnar is **responsible** for U(Ridley is mortal). God, or the laws of nature, or nothing, might be held responsible for Ridley's mortality, but not Gunnar. Consequently, concludes van Inwagen, the counterexample fails, and PPP2 is valid.

But this argument surely rests upon a mistake. The mistake is in thinking, as van Inwagen does, that U(Ridley dies) is the same state of affairs as U(Ridley is mortal). It seems clear to me that it isn't. Rather U(Ridley dies) corresponds to the state of affairs U(Ridley is shown to be mortal). And although Gunnar cannot be held responsible for U(Ridley is mortal) he can be held responsible for U(Ridley is shown to be mortal). Once this correct state of affairs is inserted into the argument above then one can see that there is nothing wrong with it. Consequently, I conclude that the counterexample is valid, and that PPP2, like PPP1, fails to vindicate the C-condition.

#### 4.ii.c Robert Heinaman's objections

Robert Heinaman objects to Frankfurt's refutation of PAP because he believes that the sense of 'could have done

otherwise' used by PAP is different to that which occurs in Frankfurt's counterexample. He states:

When the principle asserts that an agent is responsible for his action only if **he could have done otherwise**, the emphasised statement is true only if prior conditions together with the laws of nature do not (physically or naturally) necessitate the action. If prior conditions together with the laws of nature do necessitate the action, then the agent could not have done otherwise in the sense of 'could have done otherwise' relevant to the Principle of Alternate Possibilities...

This is not the sense in which Jones cannot do otherwise in the original example. For in Frankfurt's use of that expression Jones cannot do otherwise even if determinism is false and it is not the case that prior conditions together with the laws of nature necessitate the act (1986, pp.275-276)

I find this argument unconvincing. It seems to rest upon the dubious premise that because we can never do otherwise if determinism holds true that we must always be able to do otherwise if it does not hold true. But this is clearly not the case. For it implies that if the conditions sufficient for A to obtain are not deterministic that 'not A' could have obtained. For example, it implies that if Jones is indeterministically caused to push Smith off a cliff (by an indeterministic process in Jones' brain) that Smith could have done otherwise than be pushed off the cliff because prior conditions together with the laws of nature did not necessitate his being pushed off the cliff. But in this instance how **could** have Smith have done otherwise than be pushed off the cliff?

Clearly, when we say that a person could or could not have done otherwise we do not just have determinism in mind. In fact, we very rarely ever have determinism in mind when we believe that we could or could not have done otherwise than we did. Consequently, to state that PAP

only applies to determinism is not just begging the question against Frankfurt, it is surely false.

#### 4.ii.d Mark Strasser's objections

Mark Strasser accepts that Frankfurt's and others' counterexamples apply to PAP as it stands, but like van Inwagen, he believes that there are variations of PAP which are immune to such counterexamples. He states that whilst 'counterexamples to PAP might be valid if we only think in terms of the present, they will not be successful if we include past alternatives as well'. (1988, p.240) He therefore proposes the following as an amendment to PAP:

Principle of Alternate Present or Past Possibilities (PAPPP): A person is morally responsible for what he has done only if:

- a. he could have done otherwise, either presently or in the past, and
- b. he, as a reasonable agent, would have known to have done otherwise. (ibid, p.244)

Although this may overcome the problems raised by Frankfurt's counterexample, I fail to see why it should be immune to Frankfurt-style counterexamples. Imagine that Smith knows that there is a good chance that he will attack someone today (because he knows that certain atmospheric conditions make him attack people, and believes, rightly, that those conditions may occur today) but that even so he decides to go out. Unknown to him, however, Black was present and would have made Smith go out if he had decided not to do so. Consequently, even though Smith could not have done otherwise, yet he chose to go out without Black's interference, and so will be

morally responsible for his later act of attacking someone, even though he will not be able to do otherwise when he attacks them, and even though he could not have done otherwise when he chose to go out.

In fact, it will always be possible to create counterexamples to PAP variations which rely on an agent's ability to have done otherwise at some point in the past. Imagine that God creates a world in which people can potentially fulfil the C-condition. God knows that if certain choices are made that at a certain time in history Jones will make a discovery of great benefit to mankind. God, being a nice fellow, wants Jones to make this discovery and so he prepares to watch over Jones throughout his life up until the time he makes the discovery: if it appears as if Jones is going to do something which will not lead him to make the discovery God is prepared to intervene and actually make him do the right thing; he will set him on the 'right track' again. As it happens, however, Jones does everything exactly as he needs to in order to make the discovery: God does not need to interfere. Consequently, everything takes place as if God was not there. And yet he clearly was there, and would have prevented Jones from doing anything except that which Jones did do (for anything else would not have lead to the discovery.) In other words, at no time in Jones life could he do otherwise (because of God) and yet (because God did not interfere) Jones is surely as responsible for his discovery as he would have been if God had not been present? For God's presence was, in the end, as irrelevant as Black's presence in Frankfurt's original

example. I therefore conclude that the C-condition is not vindicated by Strasser's variation of PAP.

This ends my discussion of Frankfurt's refutation of the C-condition. None of the criticisms of Frankfurt's argument (or variations on it) appear to have provided any sound reason to doubt its legitimacy, and so I conclude that the C-condition is not a necessary condition of moral responsibility. Nor, therefore, can it be a necessary condition of free will, for the arguments which refute PAP as it applies to moral responsibility also refute PAP as it applies to free will. I take this further conclusion to be uncontentious following the arguments against PAP as it applies to moral responsibility. We can conceive of a situation in which Jones cannot decide otherwise, because of black, and yet, because Black does not interfere in his thought processes, his decision must count as free, for the situation is as it would have been if Black were not present.

#### 4.iii The Theory of Relativity and the C-Condition

In this section I shall argue that the theory of relativity is incompatible with the C-condition (it is therefore of particular relevance to those libertarians who are suspicious of the validity of Frankfurt's refutation of the C-condition, discussed in the last section.) I shall not be concerned with any technical details of the theory, but only with one of its well known consequences concerning the nature of time. Consequences which have profound implications for libertarian free



will.

It is generally agreed by physicists that one of the consequences of the theory of relativity is that we should abandon the common sense view of reality as being composed of three dimensions of space and a clearly distinct fourth dimension of time in favour of the view that reality is actually composed of four-dimensional spacetime (see, for example, Davies, 1983, p.124; Hawking, 1988, p.24; Penrose, 1989, p.574; Adair, 1993, pp.90-93; Stannard, 1993, p.34.) It is also agreed that the notion of four-dimensional spacetime has some profound implications for our understanding of time. For we have to accept that we are no longer dealing with a three dimensional reality evolving in time. According to the physicist Russell Stannard we have to accept:

..that our 4-D spacetime cannot evolve in time because time is not something separate from it: time is incorporated into the spacetime **itself**.

This faces us with one of the most curious features of spacetime: it never changes; it just sits there - doing nothing. All of space is there - **but so too is all of time**.

Just imagine: the whole of time - what we call the past, the present and the future - it is all there. Every instant of time is there on an equal footing with every other instant. **On an equal footing**. I mean that. (1993, p.48)

The whole idea of an absolute past, consisting of fixed events that used to exist but do so no more, and an absolute future, consisting of uncertain events that do not yet exist - all this is entirely foreign to the notion of a four-dimensional existence. I repeat: according to Relativity theory, **all of time exists**. (ibid)

This same conclusion is also described by Paul Davies:

In the..[theory of relativity]..there is no universal present, and the entire past and future of the universe are regarded as existing in an indivisible whole. The world is four dimensional (three of space, one of time),

and all events are simply there: the future does not 'happen' or 'unfold'. (1983, p.137)

The implications of this view for the C-condition should be immediately apparent. The C-condition requires that the future is open, in that it has not yet happened. It has yet to exist. This is because the C-condition requires that to have free will you must be able to control or decide your own future, to be able to choose from a number of physically possible futures. But this is simply not possible if the what we call 'the future' already exists.

According to relativity theory 'there is no universal, absolute, unambiguous distinction between past and future.' (Maxwell, 1985, p.23) But the libertarian specifically requires there to be a universal, absolute unambiguous distinction between past and future, in order to divide off the one past from the many alternative possible futures. J.R.Lucas, for one, has clearly expresses this libertarian belief:

The future is open, alterable, to some extent malleable by us. The past is closed, unalterable, part of the irrevocable record of history. (Lucas, 1986, p.126)

Consequently, if the theory of relativity is correct, the C-condition cannot be fulfilled. The future is as closed as the past.

Given that this basic argument is so simple, and so obvious, I find it odd that it has not been raised more often (to my knowledge, J.R.Lucas is the only libertarian to have explicitly considered the implications of relativity theory for free will, although Karl Popper has discussed it's implications for indeterminism, which is of

relevance to libertarian free will.)

The basic reason for this, I suppose, is that the majority of philosophers have simply been unaware of the consequences of relativity theory for libertarian free will. Either because of the fact that they do not take an active interest in science, or because they have been too engrossed with that other major theory of the twentieth century, quantum theory. They may have concluded that since quantum theory is not incompatible with the C-condition (a highly debatable conclusion in itself) that physics, and therefore science, in general is not incompatible with the C-condition. But this view ignores that fact that quantum theory is only one, albeit significant, part of modern physics. There is another part, and that is relativity theory, and it is incompatible with the C-condition. Why should we therefore not conclude that physics in general is incompatible with the C-condition?

It may be argued that this conclusion should be resisted exactly because relativity theory is incompatible with quantum theory. The objection may go something like this: quantum theory holds that events occur indeterministically, and this surely implies that the future is open? Relativity, on the other hand, holds that what we call the future already exists. Since quantum theory is backed up by a great wealth of empirical evidence, should we not suspect that there is something fundamentally wrong with relativity?

This is basically Lucas' objection. He states:

If we consider only relativistic physics, the world of black holes and temporal horizons...we are led to think in

terms of a block universe in which the future is closed, but if we consider quantum mechanics in which wave functions collapse into particles or operators when measured assume eigen-values with a probability given by their projections on the corresponding eigen-vectors, no suggestion of the future being closed can be entertained. (1986, p.130)

There are two replies to an objection of this sort. The first is to point out that even if the two theories are incompatible in this respect that relativity theory is also backed up by a great wealth of evidence:

Every week I receive manuscripts by amateur scientists intent on finding fault with Einstein's work, attempting to restore the common-sense, traditional concept of time despite almost eighty years of success during which not a single experiment has marred the flawless predictions of the theory of relativity. (Davies, 1983, p.119)

So, if the two conflict, why should it be relativity that is at fault? It could as easily be quantum theory. After all, as I stated in chapter 3, orthodox quantum theory is by no means a satisfactory theory. More realistic, of course, is the view that both theories will require minor adjustments if they conflict.

However, and this is the second reply, it is by no means certain that they do conflict in the manner stated by the objection. Quantum theory does indeed state that certain events are undetermined, but this only presupposes that the future is open in so far as it is inherently **unpredictable**. And this does not require that there are several possible alternative futures. It only presupposes that it is impossible to predict what (we call) the future will be. And this is perfectly consistent with relativity theory:

The physicist views spacetime as laid out like a map, with time extending along one side. Events are marked as points on the map - some events are linked by causal relations to prior events, others, like the decay of a radioactive nucleus, are labelled 'spontaneous'. It's all **there**,

whether the causal links are incorporated or not. (ibid, p.132)

Consequently, there is no inconsistency of the sort described in the objection, and on this point there is no reason to regard relativity suspiciously.

It almost goes without saying that on their own these 'spontaneous' events do not provide any reason to believe that relativity and libertarian free will are compatible. I mention this because Davies seems to think that they do. He says after the quote above, 'my contention that there is no past, present or future says nothing about free will or determinism at all.' (1983, p.132) But this is surely wrong? For libertarian free will does not simply require that events are undetermined, as Davies seems to think (if it did then it would be compatible with relativity). Rather, it requires that events are undetermined **and** that there are several possible alternative futures, either of which it is within the agent's power to choose. It is this which is incompatible with relativity theory.

Despite Davies assertion, it has sometimes been held that relativity theory favours a deterministic metaphysics. It is for this reason that Karl Popper, a staunch indeterminist, objected to it. Popper's objections to the implications of relativity theory were aimed at Einstein's own interpretation of these implications. In fact, he actually discussed his objections with Einstein himself, whom he referred to as Parmenides, 'since he believed in a four-dimensional block-universe, unchanging like the three-dimensional universe of Parmenides.' (Popper, 1982, p.90)

Popper compared Einstein's view of the universe with a motion picture, an analogy that Einstein apparently did not object to:

..in the eyes of God, the film was just there, and the future was there as much as the past: nothing ever happened in this world, and change was a human illusion, as was also the difference between the future and the past. (ibid)

Popper raised several objections to this view of the universe. To begin with, he argued that 'nothing in our experience of this world warranted a Parmenidean metaphysics of this kind.' (ibid) If by experience Popper meant our ordinary, everyday experience, then this does not amount to much of an objection, for, as Lewis Wolpert recently stated in his book 'The Unnatural Nature of Science', scientific ideas 'are very often outside everyday experience.' (Wolpert, 1992, p.1) The most obvious example to illustrate this, and the most apt, given Popper's staunch support of indeterminism, is quantum theory. Consequently, it is not necessarily a drawback in a scientific theory if it is at odds with our everyday experiences.

If Popper was not referring to our ordinary, everyday experience, but rather to the more specialised experience that arises with detailed experimentation then his assertion is simply wrong, for the experimental evidence is clearly consistent with a four-dimensional metaphysics (For this reason I assume that Popper did not mean 'experience' to be understood in this second way.)

Popper then argued that if the universe was assumed to be predetermined, and four-dimensional, like the film (for if we take each of its shots as representing a three-

dimensional aspect of the world, we may take the order of the sequence of the shots as the fourth dimension), that a number of unacceptable consequences would follow.

The first was that the future, being causally entailed by the past, could be viewed as contained in the past, just as the chick is contained in its egg:

Einstein's determinism made it **completely** contained in the past, in every detail. The future became, therefore, **redundant. It was superfluous.** There was little sense in watching a film all of whose shots were strictly logically entailed (in conjunction with a known theory) by its first shot. (Popper, 1982, p.91)

On the contrary, it seems to me that there is a great deal of sense in watching a film, even if all of its shots are strictly logically entailed by its first shot, if one does not know what the film contains. I do not understand why we should consider the future to be superfluous simply because it is determined, when we do not know what that future contains. This strikes me as a very weak objection.

The second consequence, according to Popper:

..was that we were bound to interpret our own human way of **experiencing change**, and the flow of time. This would have to be done, again, by using the film analogy: we experience successive shots or 'time slices' of our surrounding world, plus the successive order. But this amounts to saying that the arrow of time is subjective, and that time as we experience it is an illusion. (ibid)

In answer to this objection consider this imaginary conversation between a physicist and a sceptic:

Sceptic: You still haven't explained to me why I **feel** the flow of time.

Physicist: I'm not a neurologist. It has probably got something to do with short term memory processes.

Sceptic: You're claiming it's all in the mind - an illusion?

Physicist: You would be unwise to appeal to your feelings to attribute physical qualities to the outside world. Haven't you ever felt dizzy?

Sceptic: Of course.

Physicist: But you do not attempt to attribute your dizziness to a rotation of the universe, in spite of the fact that you feel the world is spinning round?

Sceptic: No. It's clearly an illusion.

Physicist: So, I maintain that the whirling of time is like the whirling of space - a sort of temporal dizziness - which is given a false impression of reality by our confused language, with its tense structure and meaningless phrases about the past, present and future. (Davies, 1984, p.132)

In other words, whereas the objection states that we do not feel as if we live in four dimensions, it should really state that we do not feel as we **imagine** we ought to feel if we lived in four dimensions. And stated in this way we can see it for the poor thing that it is, for as Davies says, we would be unwise to appeal to our feelings to attribute physical qualities to the outside world.

The last consequence, according to Popper, looks like a flat contradiction:

If we were experiencing successive shots of an unchanging world, then one thing, at least, was genuinely changing in this world: our conscious experience. A motion picture film, although existing now, and predetermined, has to **pass**, to **move**, through the projector (that is, relative to ourselves), in order to produce the experience, or the illusion, of temporal change. Similarly, we should have to move, relatively to the four-dimensional block universe; for the conversion of our future into our past means a change **for us**. And since we are part of the world, there would thus be change in the world - which contradicts Parmenides' view. (1982, p.91)

This problem can be overcome if one allows that change is subjective. For God, with an overall view of all of space-time, there is no change, for everything is revealed before him like a map. But for us, who actually exist in this four-dimensional universe, change does occur; it occurs when we move about through space in time.



Thus we can reconcile the fact that we experience change with the fact that, from a superior point of view, there is no change.

What other options are open to the libertarian when faced with the assertion that it is incompatible with the C-condition? He could I suppose argue that the mind occupies a purely mental realm, and so lies outside the laws of physics, and is therefore quite unaffected by the theory of relativity. But a dualist option, as I mentioned in chapter 3, has very little to recommend it, and given that this mind would still have to interact with a physical world that behaved in accordance with the theory of relativity, it is difficult to see what use it would be if it itself were not governed by this theory. It would have a freedom that could not be put to use.

One other option open to the libertarian is to hold out the hope that the theory of relativity is incomplete. At some point in the future, he could argue, just as relativity replaced Newtonian theory, so some new theory may replace the theory of relativity; a theory which is not incompatible with the C-condition. This is a slightly more serious objection.

It is more serious because it is quite likely that relativity theory will be replaced at some point in the future. However, I think that it is very unlikely that it will be replaced by a theory that differs very greatly from it. After all, the experimental predictions of Relativity theory only differ from those of Newtonian theory under certain circumstances. Furthermore, as noted above, relativity theory has a great wealth of evidence to

support it, and so any changes to the theory would have to be small, in order to preserve this empirical success. Therefore, the chances of relativity theory being replaced by a theory which fundamentally differs from it in such a way that it describes the nature of time in a more 'traditional' way, in a way that is compatible with libertarian free will, must surely be low.

It seems that the libertarian has no option but to abandon the C-condition, or to accept that his conception of free will is incompatible with the current scientific world view.

This concludes my discussion of the C-condition. I have argued that we have good reason to believe that it is superfluous to ascriptions of free will, and, that even if it is not, it is clearly inconsistent with the theory of relativity.

## Chapter Five

### The Evolution of Libertarian Free Will

The arguments of the previous two chapters suggest that we cannot fulfil one of the conditions of libertarian free will - the U-condition - and that the other condition - the C-condition - is superfluous. Nevertheless, it may be felt by the libertarian that the preceding arguments are not conclusive and that a case can still be made for libertarian free will. For this reason I now wish to consider a further objection that can be raised against libertarian free will, in order to show beyond doubt that it has very little to recommend it. I shall argue that the libertarian faces grave difficulties in explaining how libertarian free will can have evolved. Unless I specify otherwise, the terms free will and moral responsibility refer specifically to libertarian free will and moral responsibility for the remainder of this chapter.

The subject of the evolution of libertarian free will has received little coverage by either libertarians or their critics. It has certainly received no satisfactory coverage. It may be thought by libertarians that evolution does not pose a serious problem for their ideas and so is a subject not worth pursuing. But such an attitude leaves two vitally important questions unanswered:

- 1) How did free will evolve? That is, by what mechanism or process did free events evolve?

2) Why did free will evolve? That is, what conditions brought about the evolution of free will? What advantages did free will bring to the species that first had it?

Why are these questions so important? Because they must both be answered by the libertarian (and, of course, the compatibilist), or attempts made to answer them, if he wishes his theory to be empirically plausible, that is, to satisfy the second general condition of free will given in chapter 2.

Evolution is dependent upon two factors: the mutation of genes which gives rise to new chromosomes and hence to new individuals, and the natural selection of those individuals which are better adapted to their environment from those that are less well adapted. There are two points which arise from this theory which are particularly pertinent to free will. First, evolution involves very small changes in successive new species, and not giant leaps (Dawkins, 1986, ch.4.) For example, one does not go from a state of no sight to a state of sight in one mutation. Rather one goes through many small stages. Secondly, evolution is a non-random cumulative process. It involves successive random mutations building upon the advantages which prior random mutations have established, to produce the individual which is best suited to the environment in which it lives. This does not mean that natural selection is consciously leading to some end, towards some optimum individual, and that each mutation is built upon prior mutations in some purposeful fashion to reach this end. The process of natural selection is due to the totally blind forces of the environment.

Darwin wrote that if there existed any complex organs which could not possibly have been formed by numerous, successive, slight modifications, then the theory of evolution would absolutely break down (Dawkins, 1986, p.91). This argument still applies today. The brain, for example, must have evolved through numerous, successive, slight modifications. And if the brain evolved in this way, then so must the capacities that are associated with the brain, consciousness, and that which is of particular interest to this thesis, free will. It is this feature of free will which I believe to be of fundamental difficulty to libertarian free will. For the libertarian is simply unable to explain how libertarian free will could have evolved in the gradual manner that evolution requires.

Even at the outset it should be obvious that the libertarian is faced with certain unavoidable problems in discussing the evolution of free will. In order to understand **how** free will could have evolved - by what mechanism or process it evolved - it is necessary, at the very least, to have some idea of the mechanism or process by which free will arises in us now. But this, as the discussion in chapter 3 has shown, is something that the libertarian does not have. He has no clear idea of the mechanism by which free will arises. Therefore, he is, even before he begins, severely limited in the sort of conclusions that he can reach over evolution. He may be able to discuss at some length the evolution of consciousness and self-consciousness, or the evolution of purposive or intentional behaviour in animals, or the evolution of quantum indeterministic events in brain

processes - all of which may be relevant to libertarian free will - but such discussions will not touch upon the pressing problem of **how** we can be the originators of our thoughts, in the libertarian sense: how we can be free, undetermined, and yet not random, in our thoughts. Until we know exactly which brain processes are involved in free will, and the mechanism by which they operate, then any firm answers to the problem of how libertarian free will could have evolved, how the relevant brain processes could have evolved, will remain forever beyond our reach.

The libertarian may find this too extreme. All right, he may reply, the mechanism by which free will arises may be lacking at the present time, but this does not mean that libertarian free will could not have evolved. As long as it can be shown, in outline at least, that free will could have evolved through the accumulation of slight, successive modifications, then the lack of detail is no more than an irritation. It is certainly not a refutation. Especially since a similar problem arises for the compatibilist. He is limited in his answer to the problem of how free will can have evolved by his limited knowledge of the way in which consciousness, and therefore free will, arises through the complex processes of the brain.

The problem for the libertarian, though, is that it is by no means clear, even in outline, how something like libertarian free will could have developed through the accumulation of slight, successive, modifications. This is not a problem which the compatibilist shares. It is not inconceivable that indeterministic brain processes of the

sort posited by John Eccles (1986, 1989) could have evolved through slight successive modifications. One can imagine brain processes evolving that were of such a small size that they happened to be affected by Heisenberg's uncertainty principle. And that after a period of time such quantum effects came to have a significant affect upon the organism's otherwise deterministic thought processes. This much at least appears to be possible, although a bit fanciful. But it remains completely unclear how these indeterministic processes could have developed, in combination with deterministic processes, into processes which gave rise to free thoughts.

Nor is the problem simply one of lack of detail (for if it were, then the libertarian would face no greater difficulties than the compatibilist in explaining how free will can have evolved.) Rather it concerns the nature of libertarian free will itself. Consider the case of the libertarian who believes that our mental interactions, our thought processes, can be explained in terms of agent causality (ignoring the problems raised against agent causality in chapter 3.) He believes that these interactions cannot be reduced to, or explained in terms of, event causality. Yet organisms whose mental interactions are explicable in terms of agent causality must have evolved from organisms whose mental interactions are explicable purely in terms of event causality, for human evolution can be traced back to primitive individuals who not only did not possess free will, but were not even conscious. Their behaviour would have been completely explicable in terms of event causality (at

least, in principle), much like the behaviour of primitive animals is today. There would be no need to turn to agent causality. But this raises a profound puzzle. For the libertarian would have us believe that even though agent causality cannot be reduced to event causality, organisms whose minds (partly) interact in terms of agent causality somehow evolved from organisms whose minds only interacted in terms of event causality. This is not simply confusing, it is surely illogical? How could agent causality have evolved from event causality through the slight, successive modifications which evolution requires if agent causality cannot be explained in terms of event causality: if agent causality is a fundamentally different kind of causality to event causality? *With evidence of consciousness*

How can the brain have gone through a succession of slight modifications from an organ that worked through event causality to an organ that worked, at least partly, through agent causality, if agent causality is not explicable in terms of event causality? For agent causality to have evolved there must have been a series of brains, from the purely 'event causality brain' through to the 'agent causality brain', each of which only differed slightly in it's structure from the brain from which it evolved and from the brain into which it evolved. But how can we square this with the claim that agent causality is fundamentally different to event causality? If agent causality brains evolved from event causality brains then agent causality brains simply cannot be fundamentally different, for otherwise they simply could not have evolved. Consequently, if one wants to argue that



libertarian free will evolved one must accept that agent causality cannot be fundamentally different from event causality. But if one accepts this, then one must accept that libertarian free will is simply not possible, for as I argued in chapter 3, and as is agreed upon by both compatibilists and libertarians alike, libertarian free will must be something more than some combination of the random and caused events of event causality.

The libertarian is therefore faced with severe difficulties when it comes to explaining how free will can have evolved. However, he may be able to salvage something if he can explain why free will must have evolved: if he can show that there are clear advantages to our having free will. This will not make the mechanism by which free will evolved any clearer, but at least it will provide some evidence in its favour.

What then are the advantages to having free will? One possible advantage has been suggested by the libertarian Richard Swinburne:

Natural selection favoured the evolution of organisms whose purposes were produced by a non-deterministic mechanism, because of the evolutionary advantages which would be possessed by organisms who tried to defeat the predictions of predators and those who would enslave them, the forecasts of experts and commands of authority! (1986, p.259)

There are three shortcomings to this proposal. The first is the rather trivial one that in many instances humans are all too clearly predictable. If we do possess some form of indeterministic mechanism which enables us to avoid predictable behaviour then it seems, at times, that it goes oddly awry, or is simply ineffective. Or perhaps it is just that we choose to be predictable, no longer

having any need to avoid predators.

The second shortcoming is more telling. Why should early humans have evolved an indeterministic mechanism to be unpredictable when they, and many other species, would have been so complex that even if they were completely determined it would have been practically impossible, in many cases, to determine their future behaviour in any satisfactory way. They simply would not have needed any indeterministic mechanism to be practically unpredictable. They would have only needed to have something like the random-number generators that computers have, which are really only pseudo-random in that they are deterministic, in order to be practically unpredictable (Dennett, 1984, p.119). Consequently, at the complex level of human brains, it does not seem likely that random behaviour would have added anything to pseudo-random behaviour as regards unpredictability.

The third problem for the libertarian is that the compatibilist can agree that indeterminism would be advantageous but he can then simply argue that it can be incorporated within a non-libertarian metaphysics as easily as a libertarian one. For it is not indeterministic free will which provides the necessary unpredictability but indeterminism per se. The compatibilist can argue that indeterministic processes could have evolved alongside deterministic processes in such a way as to be advantageous to the organism but without the need to recourse to the mysteries of libertarian free will. He could advocate a compatibilism + indeterminism account of free will (perhaps along the lines of Kane's so-called

libertarian account, discussed in chapter 3.)

The unpredictability argument, therefore, does not provide the libertarian with a very powerful argument for the evolution of free will. How else might indeterminism be advantageous? It could enable the spontaneous occurrence of thoughts in a way which would enable the agent to consider things it would not otherwise have considered through its determined thought processes. However, this would as often as not turn out useless thoughts rather than useful ones, and, again, it can be incorporated within a compatibilist metaphysics. The libertarian could reply that the free libertarian will would be able to pick out thoughts in a useful way that would be beyond mere indeterminism, but it is unclear how this could be achieved and simply raises again the problem of the mechanism by which libertarian free will operates.

How else might libertarian free will be evolutionarily advantageous? Could it be advantageous to be able to make either of two physically possible choices, that is, to be able to fulfil the C-condition? I do not see why it should be. For as I argued in chapter 2, there is nothing about determinism that prevents an agent from being able to deliberate, and, through deliberation, from being able to make the best decisions in order to achieve his aims. Given this, then freedom from determinism of the sort advocated by the libertarian would seem to provide us with the freedom to make alternative decisions to the (good) decisions that we would have been determined to make. That is, it would provide us with the freedom to make poor decisions. And this is surely no advantage

whatsoever?

Imagine the following scenario. You are at a zoo when you see that a lion has escaped from its enclosure. You are faced with the choice of either running away or staying where you are and being caught by the lion. A person who is determined to decide that the best thing to do is to run away has no choice in this, according to the libertarian, because he has been determined to make the deliberations that lead to this decision. The libertarian, on the other hand, not being determined, has a choice. He can choose to stay and be caught by the lion. But what sort of choice is that? It clearly offers no advantage whatsoever to the agent that has it over agents that do not.

And if we consider situations in which there is no best decision then it will not matter that an agent could have made only one decision, and not another, for there would have been no advantage in being able to make another decision. Either way, the ability to fulfil the C-condition does not appear to be advantageous.

Thus far it is very difficult to see what possible advantage libertarian free will would confer upon an agent. Perhaps it is advantageous because it enables us to be morally responsible, which is itself advantageous. But if this is the case then the libertarian needs to show, firstly, why it is advantageous to be morally responsible, and secondly, why this advantage does not also hold for compatibilist accounts of moral responsibility. For if the libertarian cannot suggest what advantages it has in this respect then there will be no reason to accept a

libertarian account of moral responsibility over a more metaphysically and scientifically satisfactory account of compatibilist moral responsibility. I have to say that I do not see how this can be done. I do not see what advantage an agent with libertarian moral responsibility would have over an agent with compatibilist moral responsibility (even if libertarian moral responsibility were not open to the objection raised in chapter 4.)

The conclusion so far must be that the evolution of free will appears to present the libertarian with a number of problems. There is, though, one obvious way in which he can try to overcome them. He can argue that whilst libertarian free will and standard evolutionary theory are incompatible, libertarian free will and a somewhat modified theory of evolution are not. That is, rather than reject either free will or evolutionary theory completely, he can reject those parts of evolutionary theory which appear to conflict with free will, whilst retaining the bulk of the theory.

Here is one way in which he might do this. I have argued that evolution raises some serious problems for the libertarian because standard evolutionary theory states that evolution takes place with the accumulation of slight, successive modifications. But the libertarian can reject this standard view and adopt one which involves saltationism: that is, he can argue that evolution sometimes takes place in large jumps. He can then state that the appearance of libertarian free will was just such a jump.

This is not an option that is worthy of much

consideration, though, and I have to say that I do not know of any libertarian who has expounded it as a means to overcome the problems raised by the evolution of libertarian free will. It has been popular with certain individuals over the years but it is not a theory that has any support from professional biologists. As I said, a moments reflection is enough to highlight its shortcomings, as I shall now briefly show.

First, if large mutations did occur then new individuals who differed too greatly from their ancestors, because of the size of their mutation, would be in grave danger of being too far removed from them to breed. They would then simply die out. It is far safer, in evolutionary terms, to change slightly, so that you do not jeopardise your chances of mating with those around you. Secondly, large modifications are far more likely to be disadvantageous than small ones (Dawkins, pp.231-233). Thirdly, large modifications are simply far less likely to occur than small ones, especially where complex organs such as the brain are involved. A large modification requires a large number of random mutations compared to those required for a small modification, thus making it far more unlikely that such large modifications will occur. Furthermore, even if such mutations did take place the chances of them combining in such a way as to produce a viable modification, favourable or not, are so tiny compared to the chances of them forming one of the vast number of unviable combinations, that such an event must be considered to be near impossible. These problems worsen as the complexity of the organ increases, since even

greater numbers of mutations will be required to produce a large change. Large modifications can therefore be seen to be not only less likely than small modifications, but to be less likely to produce any advantage.

Another way in which the libertarian can adapt the standard theory of evolution is to argue that it is incomplete, or fails in some significant way. Consequently, what is already there does not necessarily need to be changed, but rather its limitations recognised. Once we appreciate this we can see that it is not surprising that one cannot see how free will can have evolved, since standard evolutionary theory is too inadequate to be able to explain it.

Critics of this sort include Richard Swinburne:

The origination of the most novel and striking features of animals (their conscious life of feeling, choice, and reason) probably lies utterly beyond the range of science.

I conclude that the process of animal evolution, apparently so regular and predictable, is yet in the respect of those all-important properties of animals (their mental life which makes them, like humans, deserving of kindness and reverence, and which makes them also interact with ourselves) not scientifically explicable. The gradual evolution of the animal soul is a mystery, likely ever to lie beyond the capacity of science to explain. (1986, p.194-195)

And John Eccles:

It is disturbing that evolutionists have largely ignored the tremendous enigma that is presented to their materialistic theory by the emergence of mentality in the animal evolution. (1989, p.176)

The answer, these two suggest, is to invoke a higher power: God. For there are clearly no problems which He cannot overcome. Eccles states:

I believe that there is a Divine Providence operating over and above the materialist happenings of biological evolution. (1989, p.235)

In a similar vein Richard Swinburne states that we should seek a personal explanation for those phenomena which apparently lie beyond evolution:

Invoking a personal explanation in this case involves invoking God, a power behind nature, who intentionally keeps the laws of nature operative. (1986, p.198)

However, invoking God not only makes things more mysterious, it re-introduces into evolution the very thing which evolution was designed to remove: supernatural explanation. Darwin had no time for such arguments. He wrote to the geologist Charles Lyell:

If I were convinced that I required such additions to the theory of natural selection, I would reject it as rubbish...I would give nothing for the theory of Natural selection, if it requires miraculous additions at any one stage of descent. (quoted in Dawkins, 1986, p.249)

Darwin is surely right in this matter. The whole point of evolution is to explain the origin and development of life, and all those features that are associated with life, without recourse to supernatural explanations. It is not only that such explanations open up a huge can of theological worms, although that is enough. It is rather that they are completely anti-evolutionary.

As I stated earlier, evolution denies that natural selection is consciously leading to some end, towards some optimum individual, and that each mutation is built upon prior mutations in some purposeful fashion to reach this end. The process of natural selection is due to the totally blind forces of the environment. But to invoke God requires that evolution is working towards some conscious end: the evolution of man. And such a view is totally at odds with evolutionary theory.



Although there are certain obvious problems confronting evolutionary theory - specifically the appearance of life, and the evolution of consciousness - I see no reason to suppose that they will not be solved within current evolutionary metaphysics or that we need to turn to a higher being for our answers. There have been many cases in the last hundred years when evolution appeared to be at a loss for an explanation, each of which were later shown to be explicable entirely in evolutionary terms. Why should we now give up evolution in favour of mysterious explanations, which are really not explanations of any kind, in order to salvage libertarian free will? I can see no reason to, especially when libertarian free will faces so many other problems, as I have shown in this chapter.

There is a final option open to the libertarian when faced with the charge that his conception of free will is inconsistent with evolutionary theory. This is to reject the theory of evolution completely. I can think of only one reply to this: that it should be clear to any reasonable person that this is not an option worthy of serious consideration. There is too much evidence in favour of evolution and too little evidence in favour of any rivals - such as so-called creationism - to entertain such a desperate notion. If we possess free will then it must have evolved, and any denial of this state of affairs cannot be taken seriously. And if we possess free will, and the arguments of this chapter are valid, then we surely cannot possess libertarian free will, for it

remains a complete mystery as to how or why such a conception of free will can have evolved.

In chapter 1 I stated that there are three main issues that need to be tackled in the free will debate. The second issue - whether we can fulfil the conditions necessary for libertarian free will - has been the subject of the last three chapters. I have argued that there are a number of philosophical and scientific arguments that suggest that they cannot be fulfilled. These lead me to conclude that we cannot possibly possess libertarian free will.

It is therefore necessary to turn to the third issue: can we reach a satisfactory compatibilist conception of free will? This will be the subject of the rest of the thesis.

## Chapter Six

### Compatibilist Free Will

In chapter 2 I argued that to be free is to be in control of one's decisions, and that to be in control of one's decisions one needs to have alternatives open to one and to be able to choose between them in a rational or at least non-wanton way. In the following chapter I shall argue that this conception of free will is compatible with our living in a universe in which there is only one physically possible future (for convenience I shall refer to this as a deterministic universe even though I argued in chapter 4 that indeterminism was not necessarily incompatible with there only being one physically possible future.)

The principle task for the compatibilist is to establish that we can have alternatives open to us of the sort that enable free decisions even if we are determined. For the libertarian's main objection to compatibilism is that if we are determined we cannot possibly have the sort of alternatives open to us that enable free decisions.

Many compatibilists, perhaps the majority of compatibilists, have argued that the sort of alternatives that are necessary for free will are what can be termed conditional alternatives (e.g. Moore, 1912; Ayer, 1954; Kenny, 1978; Lycan, 1987; Watson, 1987a; Thornton, 1989.) They state that an agent had an alternative open to him

if, had the causal circumstances been slightly different such that he would have wanted to decide otherwise, he would have decided otherwise (because if he had wanted to decide otherwise there would have been nothing to prevent him from realising that want.)

I presume that conditional accounts of free will are so popular because they do appear to explain how we can have alternatives open to us if we are determined. In chapter 2 I described the situation in the following way: the compatibilist holds that an agent is determined to make what he consider is the best possible choice he can in order to achieve some goal, G. However, he also emphasises the fact that until the agent considers all of the possible ways in which, for all he knows, he could achieve G, until he deliberates, he cannot know what this best choice is. If he believes that he could achieve G by either choosing A, B, or C, then in order to determine which of these choices is the best one to make he has to consider the desirability of each choice. He has to deliberate over which will best enable him to achieve G. Imagine that he decides that B is the most desirable, the most likely to enable him to achieve G, and so he chooses B. According to the conditional compatibilist's account of free will, although the agent may accept that he was determined to choose B, he can still assert that he had the sort of alternatives open to him that enable him to make free decisions. This is because it is true, despite determinism, that if he had decided through his deliberation that A was the best choice to make, he would have chosen A, and if he had decided that C was the best

choice to make, he would have chosen C. Thus even though he was determined, he had alternatives open to him because certain relevant counterfactual conditionals are true: if he had wanted to do X, Y or Z, he would have done X, Y or Z.

However, despite this initial plausibility it seems to me that conditional accounts of free will are open to a serious objection. This is that we can conceive of individuals who do not have conditional alternatives open to them but who we would still wish to regard as free. (There have been other objections made against conditional accounts of free will, most notably those that centre on the conditionalist's claim that 'he could have decided otherwise' is equivalent to 'he would have decided otherwise if he had wanted to', but I think that these can be overcome.)

The objection arises from Frankfurt's argument against the C-condition discussed in chapter 4. For Frankfurt's argument applies as much to conditional alternatives as it does to physical ones: conditional alternatives are demonstrably superfluous to ascriptions of free will.

Consider the argument once again. Black wants Jones to murder Smith and so brings it about that Jones cannot decide anything but to murder Smith - if Jones decides to murder Smith Black will not interfere; if Jones appears as though he may not decide to murder Smith then Black will intervene and make sure that Jones does decide to murder Smith. Either way, Jones cannot help but decide to murder Smith. In the actual event, Jones decides to murder Smith,

and so Black does not need to intervene. Therefore, everything occurs as it would have done if Black had not been present, indeed, as if he had not even existed. Consequently, if we would wish to regard Jones' decision as free without Black's presence, then we should regard it as free even when Black is present, but does not interfere.

Now we come to the crux of the problem. The conditionalist states that an agent has an alternative open to him if he would have decided otherwise if he had wanted to. So Jones has an alternative open to him if he would have decided not to murder Smith if he had wanted to. Yet because of Black Jones does not have this alternative: if Jones had wanted to decide not to murder Smith then Black would have intervened and prevented him from so deciding. Therefore Jones clearly does not have a conditional alternative open to him: he would not have decided otherwise, even if he had wanted to. Yet we would surely still wish to regard Jones' decision as free, for it was the decision he would have made if Black had not been present. Consequently, conditional alternatives, like the physical alternatives of the libertarian, can be of no relevance to an agent's freedom.

The failure of conditional alternatives to provide us with the sort of alternatives that are required for free will may be regarded as a problem for the compatibilist but they are not his only option. I argued in chapter 2 that the compatibilist can explain our personal experiences of freedom, and the practice of deliberation, in terms of epistemic alternatives. He can argue that we

feel as if we can do one thing or another when we make a decision, and so deliberate about our future, because the future is unknown.

This prompts one to ask whether we could develop an account of free will based upon epistemic alternatives. Unfortunately it would seem not, since epistemic alternatives alone appear to be no guarantee of freedom. There are certain individuals who have epistemic alternatives open to them, but who we would not wish to describe as free. This objection to epistemic accounts of free will has recently been raised by Susan Wolf, who argues that the genuine freedom provided by what she terms the psychological ability to decide otherwise can clearly be distinguished from the apparent freedom provided by the epistemic (or epistemological) ability to decide otherwise. She provides the following example to show the significance of the distinction:

Tony, the son of a Mafia don, has reached a turning point in his life. He must decide whether to follow in his father's footsteps as a leader in organized crime, or instead to break completely from his family to lead a life as an honest schoolteacher. He believes that it is in his power to choose either path...Reasoning that if he doesn't take over his father's position his more ruthless and less intelligent brother will, and that his leaving will break his mother's heart, he decides to remain. Tony, and presumably everyone else, believes that Tony could have chosen otherwise. But this is an illusion. In fact, Tony's fear of his father's wrath - a consequence of his possible decision to become a school teacher to which Tony consciously accorded some, but limited weight - was greater than Tony, or anyone else, knew, and had the unconscious effect of shaping his deliberative processes, making him attach inordinate weight to certain factors and inordinately little to others so as to produce a rationalised decision that his uncontrollable unconscious fears compelled him to reach.

Since, for all Tony knew, he was able to choose otherwise, we may imagine that Tony regarded himself as fully responsible for his decision. Nonetheless, it seems to me that Tony was not fully responsible, and this because he wasn't **really** able to choose anything else, in a sense that is relevant to assessments of freedom and

responsibility. Note, however, that this sense is a sense that is captured at the psychological level of explanation. For, despite what Tony and everyone else thinks, it is not really compatible with Tony's psychological history and all the psychological laws that apply to him that he choose to break with his family...there is something - namely, fear of his father's wrath - that prevents Tony from exercising his capacities in that direction. Although Tony is, if you like, epistemologically free to choose otherwise, he is not psychologically free to chose otherwise, and therefore he is not fully responsible.

This example will, I hope, bring out the fact that psychological freedom is a good deal more substantive than epistemological freedom. Epistemological freedom, we might say, is only apparent freedom. (1990, pp.112-113)

Clearly, then, epistemic alternatives are no guarantee of freedom. But what about psychological alternatives? Do they provide us with genuine freedom, as Wolf claims? I do not believe that they do. Consider, for example, Wolf's explanation as to how we can be physically determined and psychologically free. She asks us to imagine that God has created a physically determined world in which there are psychologically free agents, such that:

..when these agents chose what to do, it would have been equally compatible with their psychological histories in conjunction with all the psychological laws applying to them that they had chosen something else. (1990, p.103)

Presumably, the psychological laws in question are general enough to fail to determine what the person's choice will be, and yet specific enough to discount agents like Tony from the category of free agents. But why should we assume that psychological laws are this general? Why should we not assume that there are psychological laws that are as specific as physical laws and therefore that if we are physically determined that we must be psychologically determined as well. As Roy Weatherford says:

If the materialists are right, and human beings are only a body, its states, and its actions, so that no "spiritual"



substances exist, then psychological events are merely a sub-class of physical events, and if all the events in the larger class are determined, it is even more certain that all the events in the smaller class are. (1991, p.9)

Even though we do not know at present what these specific psychological laws are, we should not assume that they do not exist. In which case it will not hold that after an agent has made a choice it would have been equally compatible with his psychological history in conjunction with all the psychological laws applying to him that he had chosen something else. Unless, that is, we wish to hold that an agent's psychological history and the various psychological laws applying to him are irrelevant to his choice and so it does not matter that they remain unchanged, or, that the psychological laws in question are indeterministic.

The first option is quite unacceptable, of course. How is one to make sense of a conception of psychological freedom in which the person's psychological history and the various psychological laws applying to him are irrelevant? The second option is no better. To begin with it assumes that we can make sense of the idea of an indeterministic psychological law, and that we could have such indeterministic laws within a physically determined universe. This is no small assumption. More serious, though, is the implication that the fate of one's choice depends on an indeterministic psychological law (or laws). For this would equate freedom with randomness. In which case it would appear that psychological freedom suffers from the same problem that compatibilists claim arises for libertarian freedom with respect to indeterministic

physical events. It is surely a step backward for the compatibilist if his account of free will is subject to the same objection?

The only option open to the exponent of psychological alternatives is to deny that physical determinism entails psychological determinism. And this is possible, according to some, if one holds that there are no law-like correlations between the physical realm and the psychological realm. Such a view is held by Kenny. He argues that psychological determinism is not entailed by physiological determinism - the idea that all human activity is determined via neurophysiological states of the brain and central nervous system - because there are not law-like correlations between the two:

..physiological determinism would entail psychological determinism only if physiological events of a particular kind were correlated in a regular and law-like manner with psychological conditions of a particular kind. But there is no reason to believe that physiological determinism must involve such regular correlations. It may be, for all we know, that for each individual case in which a human being can choose whether to do X or not to do X there is a difference between the state of the brain and of the central nervous system which goes with wanting to do X, and the state which goes with not wanting to do X; and this could well be the case without there being any general laws linking physiological states of a particular kind with psychological states of a particular kind. If this is so, there is no reason why physiological determinism should lead to psychological determinism, or why predictability at a physiological level should involve predictability at a psychological level. (1978, p.32)

This view has been recently endorsed by Mark Thornton, who states:

Physiological determinism would imply psychological determinism only if there were laws relating (or a 'law-like' relation between) the physiological level and the psychological level. There might conceivably be such a relationship; but there certainly need not be. (1989, pp.133-134)

But it seems to me far more plausible that there is such a relationship than that there is not. For otherwise we would have to admit that whilst there appear to be law-like correlations between the physiological and the biochemical, and between the biochemical and the physical, there are no such correlations between the psychological and the physiological. We would have to admit that the psychological realm is in some significant sense apart from the physical realm. But for anyone who wishes to take a firmly materialist or physicalist line (which I presume includes compatibilists) this conclusion must be unacceptable. For it asserts that although the mental is a part of the physical world it is not correlated with the rest of this world in a law-like way.

It may be pointed out that such a view is not inconsistent with the 'anomalous monism' of Donald Davidson (1979). Davidson argued that there are no psycho-physical laws connecting mental events with other physical events. Consequently, one can have an anomalous explanation of the mind (anomalous in that it does not fall under a law) that is consistent with a materialist or physicalist metaphysics. But such a theory of the mind is subject to two serious objections. Firstly, it seems quite clear that there can be psycho-physical laws of the kind required, as Crane and Mellor have recently argued (1990, pp.196-199). Secondly, if there are no laws connecting the mental to the physical then it seems that the mental simply becomes dispensable when one wishes to explain people's actions. In other words, anomalous monism, and closely related views of the mind, lead to

epiphenomenalism, and this is to be objected to since 'epiphenomenalism is as unbelievable as very nearly all of us take it to be.' (Honderich, 1991, pp.166-167, also 1988a, p.91, pp.96-99)

It is therefore my conclusion that one cannot reasonably hold that after an agent has made a choice it would have been equally compatible with his psychological history in conjunction with all the psychological laws applying to him that he had chosen something else. We do not have such psychological alternatives open to us.

So far I have discussed three options that are open to the compatibilist who wishes to determine what sort of alternatives are open to a person in a deterministic world: epistemic alternatives, psychological alternatives and conditional alternatives. And I have argued that there are grounds for rejecting all three. Where does one go from here?

The answer should be clearer once we bear in mind the following two points. The first is that it seems clear that epistemic alternatives must be involved in compatibilist free will to some extent, since they provide the only explanation as to how we can make sense of deliberation in a deterministic universe (as I argued in chapter 2.)

The second, which only becomes obvious in the light of the first point, is that psychological and conditional alternatives are manifestly unacceptable in a way that epistemic alternatives are not. Psychological alternatives require that one adopt an epiphenomenalist account of the mind, and conditional alternatives are simply superfluous

to ascriptions of free will. But epistemic alternatives do not require one to adopt an epiphenomenalist account of the mind, nor are they superfluous. They are clearly relevant to compatibilist free will to some extent - as can be seen from their importance to deliberation. Unfortunately, as Wolf's example shows, they appear to lead to an account of free will that is unavoidably subjective. But is this an unavoidable problem or could we modify the basic epistemic account of free will in order to overcome this subjectivity?

The answer is to be found in the work of Richard Double (1991.) Double, after analysing a number of recent compatibilist accounts of free will (Frankfurt, 1971; Watson, 1975; Levin 1979; Dennett, 1984) has formulated what he terms the autonomy variable account of free will. He argues that there are five variables that must be met by an agent if it is to be free: self knowledge, reasonability, intelligence, efficacy and unity. He states that S's choice c is free just in case:

1. S knows the nature of S's beliefs, desires and other mental states that bring about c (self knowledge).
2. S desires to perform a critical and nondogmatic evaluation of c and the mental states that bring about c in cases where such evaluation is appropriate (reasonability).
3. To the extent that reasoning is appropriate, S's reasoning concerning c and those other states meets normative standards of intellectual skill (intelligence).
4. S possesses the power, at each step in the decision-making process to produce subsequent deliberations in accordance with 1, 2, and 3 (efficacy).
5. There is a single agent to whom variables (1) through (4) apply (unity). (Double, 1991, p.48)

Although Double's account is admirably comprehensive, there is one problem that it conspicuously fails to address: the nature of the alternatives that are open to

an agent who fulfils the five autonomy variables. Double appears to hint in certain passages of his book that he believes that the compatibilist should favour conditional alternatives (1991, p.13, pp.219-220), but at no time does he go so far as to actually discuss the nature of the alternatives that are available to an agent who fulfils the autonomy variables. Given that the libertarian is entitled to know exactly what alternatives the compatibilist is suggesting are open to us, and considering the problems associated with conditional alternatives, this is hardly satisfactory. However, this deficiency can be overcome when one realises that in order to fulfil the autonomy variable account an agent only needs to have epistemic alternatives open to him. In other words, Double's account should be regarded as an epistemic account of free will.

As to the problem of the subjectivity of epistemic accounts this can be overcome by the first autonomy variable, self knowledge. Double states that if there are any hidden psychological constraints influencing our decisions - fears, desires or beliefs that are hidden from our conscious awareness - then we do not have the requisite self-knowledge of our psychological states:

Suppose that you are about to make a deliberate choice to do A because you wish to do A, and you think that you have this wish because it is supported by considerations B and C. Suppose further that, as the reasonability variable requires, you go on to evaluate B and C. It is important that B and C are in fact causally responsible for that wish. If the cause of your wish to do A is actually an entirely unrelated factor Z, then considering B and C will be merely an intellectual exercise, a gear not attached to anything. Thus, as long as Z's role is out of your awareness, you do not satisfy the account's requirements for free will, since you lack rational control over your decision to do A. (1991, p.40)

In the autonomy variable account of free will we therefore have an epistemic account of free will that is not subjective. It states that an agent who has epistemic alternatives open to him will not be free if he does not satisfy the self-knowledge variable. This is exactly the case with Tony, discussed earlier. Tony had epistemic alternatives open to him, because, as far as he knew, he could do either X or Y. However, because the reason behind his decision (his fear of his father) was not known to him, because his self-knowledge was inadequate, his decision was not free. The reasons which he believed were behind his choice were not the real reasons behind his choice.

We can therefore conclude: 1) that an agent is capable of free decisions if he can rationally evaluate the best decision to make out of the epistemic alternatives open to him; and 2) that in order to rationally evaluate the best decision to make the agent needs to fulfil the five autonomy variables. He needs to have (1) sufficient self-knowledge such that the reasons he believes are behind his decisions really are the reasons behind his decisions; (2) sufficient reasonability to be able to evaluate those reasons impartially; (3) sufficient intelligence to be able to realise (1) and (2); (4) sufficient efficacy to be able to resist non-rational factors, such as threats and bribes from the external world, as well as non-rational motivations from within the cranium; and (5) sufficient unity such that one part of his self cannot disagree with, or ignore, another part of his self and thus prevent him from being able to

successfully realise variables (1) to (4). (See Double, 1989, pp.168-170; and 1991, pp.38-49 for more detailed discussion of the autonomy variable account of free will.)

One final point concerning the autonomy variable account. Although I think that it must be a requirement of freedom that an agent is capable of making rational decisions, it seems to me that it does not follow from this that in order to be free an agent's decisions must always be rational. To return to a point I raised in chapter 2, to be free it is enough that the agent's decisions are not made wantonly; that is, without any purpose in mind. In fact, I think that even quite irrational decisions can be considered to be free if they are not made wantonly and if they are made by an agent who is potentially capable of rational decisions. In this I differ from Double who argues for quite a strong rationality component. In my opinion, though, one can be in control of one's decisions even if those decisions are on occasion quite irrational; for one may not be aware that they are, or one may have overriding motivations, such as love, hunger, or greed, which cause one to ignore the rationality of one's decisions.

If my decisions are wanton, however, if they are made with no purpose in mind, indeed, if they have no purpose, then they are not free. They may be made voluntarily, of course, such as when I decide to walk through Hampstead Heath and happen to follow one path rather than another. My decision is wanton in that I did not for one moment consciously think about the path that I was taking through the Heath, and yet I was not forced to take the path that



I did. Consequently, pace traditional compatibilism, it is not necessarily the case that voluntary decisions are equivalent to free decisions. In other words, the compatibilist should distinguish between free decisions, unfree but voluntary decisions, and forced decisions.

This basically concludes my exposition of compatibilist free will. However, there are a number of objections that may arise against such an account of free will and I wish to consider them before I continue.

Firstly, it may be objected that under this account of free will an agent like Tony has the same alternatives open to him as free agents do, since Tony is as ignorant of his future as ordinary free agents are, and this is all that is required in order to have epistemic alternatives open to one. But surely we do not wish to regard agents like Tony as having the same alternatives open to them as free agents have, or even as having alternatives at all?

This objection arises because the libertarian equates having alternatives with having a physically open future. Since agents like Tony do not have physically open futures, whereas (the libertarian argues) free agents do, they clearly do not have the same alternatives as free agents, if they have any alternatives at all. Consequently, the libertarian will naturally object to any account of alternatives that does not preserve this distinction.

My reply is to state that if one rejects the idea that one requires a physically open future to have alternatives open to one, as the compatibilist does, then one can see that there is nothing odd about regarding

agents like Tony as having the same alternatives open to them as us. If the future is not physically open then neither agents like Tony, or free agents, have alternatives open to them in the libertarian sense. And since the only other valid alternatives that are open to us - epistemic alternatives - are as available to agents like Tony as they are to free agents, then we should not find it in the least unusual that such subjects have the same alternatives open to them that free agents do. This will only seem unusual to the libertarian.

The point to stress is that only free agents are able to make rational use of the epistemic alternatives that are open to them. This is because free agents will be aware of the mental states that lie behind their decisions (they will have sufficient self-knowledge) whereas Tony was not be aware of his (at least, not with regard to those decisions that were influenced by his subconscious fear of his father.) Consequently, one should not make the mistake of assuming, just because they have the same alternatives open to them, that agents like Tony are in a comparable position to free agents; something which the objector may be hinting at in this objection.

A second objection arises from the fact that epistemic alternatives are unavoidably related to the subject's knowledge of their own future. On the one hand, if you are too ignorant of your future you may come to believe that you do not have any alternatives open to you, and so you will not be free. But on the other hand, if you are too knowledgeable of your future then you will not believe that you have any alternatives open to you because

you will appreciate the path that the determined universe, including yourself, will follow, and so you will see what you will be determined to do (c.f. Laplace's supreme 'Intelligence'.) So your having epistemic alternatives open to you is dependent upon your being neither too ignorant nor too intelligent.

Neither of these objections is significant. The problem of over-ignorance is not really a problem, since this applies to any conception of alternatives. Even in a libertarian universe, if we are too ignorant we may not appreciate the alternatives that may be there. As noted above, free will clearly requires a certain amount of intelligence: in this instance, the intelligence to appreciate that one may be determined to do X, Y or Z, and so consider the possibility that either may occur, rather than simply assuming that only X will occur, when Y or Z might occur instead.

The problem of having too much knowledge presents no more of a difficulty. To begin with, no human even remotely possess such knowledge of their future, nor is it likely that they ever will. Human life is simply too complicated for its future to be predicted. Secondly, and more importantly, no individual can have complete knowledge of his or her own states and so completely predict what he or she will be able to do (Popper, 1951; MacKay, 1960.) Consequently, the likelihood of a person gaining so much knowledge that they cease to have epistemic alternatives is not only unlikely, it is impossible.

Finally, it may be objected, especially by the

libertarian, that epistemic alternatives are not genuine alternatives: merely believing that we have alternatives open to us is not the same as actually having alternatives open to us. (This objection has already been considered, and answered, in one form, when made by Wolf, but the libertarian's objection requires a different response.)

If by 'actually having alternatives' the libertarian means having a physically open future, then of course he is right, but the compatibilist is not trying to argue that we have alternatives open to us in the libertarian's sense, since he accepts that if we are determined we cannot physically do otherwise than we do. In fact he denies that we need, let alone have, the 'actual' alternatives that the libertarian describes. The compatibilist only wishes to argue that regardless of determinism we have the only alternatives open to us that are necessary to be in control of our decisions, and so to have free will. Furthermore, such alternatives explain how we can meaningfully deliberate, and how we can be held morally responsible for our actions (chapter 7.) From that point of view epistemic alternatives, with suitable rationality additions, are genuine alternatives. The fact that they do not equate to the alternatives of the libertarian is therefore irrelevant.

This objection is in fact one version of what must be the most popular objection raised against compatibilist free will: that it does not provide us with genuine free will; although compatibilist accounts may satisfy various 'freedom criteria' still they do not provide us with real alternatives, or genuine control over our lives. Real

freedom is not compatible with determinism.

Although this line of argument is popular with libertarians it seems to me that it completely begs the question. Peter van Inwagen, for example, without any justification, states that the only alternatives that we are interested in as regards free will are physical ones, and that since these are clearly incompatible with determinism, that free will must be incompatible with determinism as well (1983, pp.8-10.) Any other alternatives are simply not genuine alternatives. Hopefully, the discussion of chapter 2 has shown that to presuppose that the only valid notion of free will is one that posits physical alternatives is to beg the question. Our ordinary, everyday, 'prephilosophical' notion of free will arises from our everyday experiences of the behaviour of others and our own personal feelings, and although people invariably interpret these experiences in a libertarian way, we can interpret them in a compatibilist way. Consequently, presupposing that the only valid notion of free will is one that posits physical alternatives simply begs the question, for it is not.

Many libertarians criticise compatibilism by offering a string of arguments, if they can be called that, that appeal to our intuitions as to what it would be like to be determined. It is held that without physical alternatives we would be 'imprisoned' by determinism, or at the mercy of the laws of physics or the environment and our genetic history. In a well known passage Isaiah Berlin states that determinism:

for all that its chains are decked out with flowers, and despite its parade of noble stoicism and the splendour and

vastness of its cosmic design, nevertheless represents the universe as a prison. (1954, p.68)

Smith and Jones offer the following antidote to the fear expressed by Berlin:

Here is Jill sunning herself on the beach of an idyllic Greek island. She has decided after some deliberation to blow a small inheritance on the holiday of a lifetime...Happily she turns over to tan her other side, decides to have another glass of wine, reaches for her book...and she is in prison? (1986, p.267)

The libertarian may reply that a happy prisoner is a prisoner none the less. But, the compatibilist can reply, if the prisoner is totally unaware that she is imprisoned, then what possible difference can the prison make? Is the ignorant prisoner still a prisoner? A further point, made by Dennett, is that 'a jail without a jailer is not a jail.' (1984, p.8) Unless we wish to believe in a 'God of determinism', or some other suitable deterministic jailer, then why should we regard determinism to be imprisoning? The libertarian may reply that a prison is a prison, regardless of whether there is anyone to do the imprisoning, or regardless of whether one is ignorant of the fact that one is imprisoned. But given that there are no sound reasons to doubt that agents can reason, deliberate, argue, discuss, debate, laugh, cry, and generally have a damn good time even if they are determined, why should one regard a deterministic universe to be imprisoning compared to an indeterministic one?

In the end the libertarian only has one answer: one cannot originate one's decisions, or decide otherwise, if one is determined. But to assume that these are requirements of free will is, as we saw above, to beg the question. There are no sound reasons to believe that

libertarian free will is genuine freedom and that compatibilist free will is not. There are only bad ones: reasons that rely upon scaremongering, mistaken intuition, and inadequate examination of the real nature of free will. The libertarian cannot claim to offer the genuine conception of free will, any more than the compatibilist can (even though both have made this claim as Honderich, 1988b, has recently argued.) However, unlike the libertarian, with his mysterious U- and C-conditions, the compatibilist is able to provide an acceptable conception of free will that we are able to fulfil. Furthermore, as I shall show in the next chapter, exactly the same applies to compatibilist moral responsibility.

## Chapter Seven

### Compatibilist Moral Responsibility

I stated in chapter 2 that it is a generally held belief that humans can be held morally responsible for their actions, and that this is possible only because they have free will. In this chapter I wish to show how the account of free will expounded in the last chapter enables agents who are free to be morally responsible.

What is moral responsibility? There are a variety of answers to this question, some of which are not very satisfactory. A number of philosophers define moral responsibility in terms of certain 'reactive' attitudes and practices such as praise, blame, punishment and reward:

Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i.e. of all we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation and justice. (P.F.Strawson, 1962, p.78)

A person is a morally responsible agent when he is an **appropriate candidate** for the reactive attitudes and for such attitudes as praise and blame and punishment and reward. (Fischer, 1986, p.12)

..to be capable of being truly responsible for one's actions is to be capable of being truly deserving of praise and blame. (G.Strawson, 1986, p.1)

But this represents the relationship between such attitudes and practices and moral responsibility in completely the wrong way. One should not argue that an agent is morally responsible **because** he is an appropriate subject for such attitudes and practices, but rather that



he is an appropriate subject for such attitudes and practices **because** he is morally responsible.

I agree with J.R. Lucas that when discussing moral responsibility the obvious place to start is to look at the original meaning of the term responsible:

Etymologically, to be responsible is to be answerable - it comes from the Latin respondeo, I answer, or the French repondre, as in RSVP. I can equally well say I am answerable for an action or accountable for it. And if I am to answer, I must have a question; the question is 'Why did you do it?' and in answering that question, I give an account..of my action. So the central core of the concept of responsibility is that I can be asked the question 'Why did you do it?' and be obliged to give an answer. (Lucas, 1993, p.5)

To be morally responsible, therefore, is to answerable for the moral consequences of one's acts. (Strictly speaking I should add 'or omissions' after acts, but this would only unnecessarily complicate the issue. It should be born in mind, though, that someone can be as responsible for what they do not do as for what they do.) Given this, it might be thought that having reached a satisfactory compatibilist account of free will the issue of a satisfactory account of compatibilist moral responsibility would be something of a formality. After all, if we make our decisions freely, and if we have an adequate moral understanding such that we can appreciate the moral consequences of our decisions, and the acts which result from then, it would seem to follow that we must be morally responsible. For if all these conditions are fulfilled then surely we have to accept that we, and we alone, are answerable for the moral consequences of our acts. If an act arose from our own free decision - if it was within our control - and if we were fully aware of the

morality of the act - if we did not commit it in moral ignorance - then it is clearly we alone who have to account for the moral consequences of the act. When asked 'Why did you do it?' we cannot (rightly) answer 'because I was forced to', or 'because I did not realise it was an unacceptable act.' We have to answer in terms of the moral (and non-moral) reasons that caused us to act. Thus it is we alone who are morally responsible for the act. Put simply, if our decisions are free, as I have argued they are, and moral, as many of them are, then they must also be ones for which we are morally responsible, in spite of determinism. What more is there to say?

Unfortunately, it is not this simple. For although the opponents of compatibilism may admit that there is a certain sense in which our decisions can be free, despite determinism, they do not accept that it is the sort of freedom that enables us to be morally responsible. They may accept that we can be in control of our decisions in a desirable and meaningful way, despite determinism; that there is a valid difference between the decisions of a mentally ill person, or a hypnotised person, and the decisions of a free person, even in a determined universe, and that we can in a sense label the decisions of the former unfree and the latter free. But they will stop short of allowing that this freedom is sufficient, with the other conditions mentioned, to enable us to be morally responsible. It is therefore necessary to consider these objections, and show why they have no force.

## 7.i Some Objections to Compatibilist Moral Responsibility

Some objections are prompted by arguments which can be instantly dismissed. For example, it may be argued that we cannot be morally responsible in a deterministic universe because we cannot avoid doing what we do. In such a universe there is only one possible course for events to follow, and so it would be unjust to hold people morally responsible for their acts when they could not avoid doing them. In answer to the question 'Why did you do it?' one can rightly say that one could not avoid doing it since one was determined to do it.

It should be clear that this objection is quite unacceptable, as it appeals to the C-condition for moral responsibility, which was rejected as superfluous in chapter 4. We do not require that agents could have done otherwise in order to hold them morally responsible, and this provides no reason to reject a compatibilist conception of moral responsibility. When a determined agent considers whether to carry out an act or not there may be a number of reasons for carrying out that act, and a number of reasons for not carrying it out, but the fact that he will be determined to do whatever he does do will not be amongst them. Consequently, for such an agent to bring this up as an excuse after the act is quite wrong, for it did not figure amongst his reasons for acting in the first place. Even though he was determined, he acted for various reasons, and thus he can rightly answer the question 'Why did you do it?' - he can answer in terms of those reasons.

It may also be argued that compatibilist moral responsibility is unacceptable because it violates the commonly held belief that the responsibility for our acts rests with the cause of our acts. That is, if we are the ultimate cause of an act of ours then we are responsible for it, but if we are not the ultimate cause of an act of ours, if the cause lies outside us (if our act was caused by a hypnotist, say, or a strong wind), then we are not responsible for it.

But this seemingly common-sense condition is in fact the U-condition for moral responsibility, which was shown in chapter 3 to be logically impossible. So this objection must be rejected as well, for it would be quite ridiculous to object to an account of moral responsibility simply because it failed to fulfil a logically impossible condition.

But could it be that moral responsibility requires the U-condition, even though it is logically impossible? That is, does the concept of moral responsibility require a logically impossible condition to be fulfilled, thereby making moral responsibility itself logically impossible? It could be the case, for Martha Klein (1990) has recently argued that the compatibilist, as much as the libertarian, must be committed to the U-condition for moral responsibility. And if this is the case, then it appears that he is indeed committed to a logically impossible condition. (Klein, it will be remembered from chapter 3, believes that the U-condition is logically possible, and so the argument I am developing is not one which she herself would wish to develop, but it is a consequence of

her argument that we are committed to the U-condition, and of Strawson's argument that the U-condition is logically impossible.)

Klein argues that the compatibilist holds someone to be morally responsible when they fulfil the M-condition, that is, when their act results from a morally reprehensible state of mind (1990, pp.11-12). She suggests as an illustration of such a state of mind 'a preparedness to do what the agent knows to be wrong for no reason which would excuse his behaviour.' (ibid, p.27) But the M-condition alone is inadequate to explain why we hold people morally responsible, for there are certain people whose morally reprehensible states of mind are abnormally caused (by hypnosis or psychoses, for example), and who we would not wish to hold morally responsible. The compatibilist therefore needs to supplement the M-condition with an additional condition, or conditions, to overcome this problem. But this is where he runs into problems, according to Klein, for the only way in which the compatibilist can successfully supplement the M-condition is with the U-condition:

The argument is this: our reactions to imaginary problem cases, which are used by philosophers to test our intuitions about moral responsibility, are such that the only satisfactory explanation for them is that we are implicitly committed to a U-condition. Such imaginary cases are those in which an agent's fulfilment of the M-condition can be attributed to his having a brain tumour or to his having had certain states of mind produced by brainwashing, hypnosis, or malevolent supernatural agents. We would be inclined to say that the victims of such implantations, brain tumours, etc. are not blameworthy and the incompatibilist would say that this response is strong evidence of our being committed to the belief that when an agent's decisions can be traced to causes for which he is not responsible, he ought not to be blamed for what he does as the result of those decisions. It is only such a belief, he would argue, which can be said to underlie and justify these responses. (ibid, p.70)

If this argument is valid, and given that the U-condition is logically impossible, then it would seem to follow that moral responsibility is logically impossible as well. This is certainly not a satisfactory conclusion as far as the compatibilist is concerned.

However, the argument is not valid, and so such a conclusion is not warranted. For the compatibilist can quite easily differentiate between those agents who are morally responsible and those who are not without recourse to the U-condition, despite Klein's claim to the contrary (ibid, p.73).

We can differentiate between those agents who have a morally reprehensible state of mind and who we would wish to hold morally responsible and those agents who have a morally reprehensible state of mind and who we would not wish to hold morally responsible by reference to the decisions which lead to their acts. The former agents are capable of free decisions whereas the latter are not. This is the only justification we need in order to differentiate between the two sorts of agents. And as we saw in chapter 6 we do not need to refer to the U-condition in order to explain whether an agent's decisions are free or not. Nor does such an explanation depend upon, or reduce to, the U-condition. Consequently, fulfilment of the U-condition should not be regarded as a requirement for moral responsibility.

## 7.ii Moral Responsibility and Reactive attitudes

There is one further objection to compatibilist moral responsibility, one which requires more consideration than the objections so far considered. It arises from what is generally taken to be a universal belief: that a morally responsible agent is an appropriate subject for certain attitudes and practices, such as praise or blame. The objection is as follows. If you are morally responsible you deserve to be praised for your moral acts and blamed for your immoral acts. This is agreed by both compatibilists and libertarians. But if all our actions are determined, say by the laws of nature and events in the distant past, then how can we really **deserve** to be blamed or praised for those actions? If my robbing the bank was determined, then I could not have done anything else but rob it, so how can I deserve to be blamed for this, when I could not have done anything else? The only conclusion is that if we are determined we do not deserve to be praised or blamed, and a part of what it is to be morally responsible cannot be achieved.

A slightly different objection, but along the same lines, is provided in more detail by Kane. Kane imagines himself to be the father of a murdered girl, attending the trial of her young murderer. He is filled with a great deal of anger and resentment towards the boy:

But as I sit in the courtroom listening to the testimony about the boy's past, a surprising thing happens. My resentment against the boy decreases as I learn more about the environmental factors influencing his character and motives. He is a mean and calculating young man, to be sure, and there is no doubt in my mind that the rape and murder were premeditated...What decreases my resentment, however, is the story of how he came to have the mean

character and perverse motives he did have, a story of poverty, parental neglect, bad role models, and so on.

To the extent that I come to believe the young man's character and motives were determined by his heredity and environment, my resentment against him as a responsible individual decreases. At first, my feelings are directed toward the parents, then toward the society which created such a cultural environment. But if I believe the characters and motives of everyone involved were determined, these feelings might shift to God, or the universe, or Fate. (1985, pp.180-181)

Kane admits that it may not be possible for his feelings of resentment and anger to be transferred to the universe or to fate, and that they will more probably be transformed, into bitterness and frustration, for example, but either way it does not seem that one is justified in feeling resentment or anger towards the boy if his actions were determined.

These sorts of objections to compatibilism are quite common, as are replies, although sometimes the replies provided by compatibilists are quite unsatisfactory. William Lycan, for example, raises the following objection to compatibilist accounts of moral responsibility:

..if the distinction between free actions and actions for which I am not responsible is just a difference between two kinds of causes, between causal pathways of two different shapes, why does it or should it have the enormous moral significance we attach to it? (When a causal chain bends one way, we pat you on the head and give you a sandwich; when it bends the other way we throw you in jail.) Is this not completely irrational? (1987, p.118)

He provides the following reply:

I grant that this seeming arbitrariness is troublesome. But notice that it is by no means confined to the free-will issue. Distinctions between causal pathways of different shapes pervade our moral life. A forged check is just a check with the wrong sort of etiology. A counterfeit bill is just one that came from a printing press on the wrong side of the tracks. A forged painting lacks the right provenance; etc. Yet we make sharp evaluative distinctions in each case. Maybe this is irrational; certainly it can be made to sound suspicious. But it is entirely natural and commonplace. (ibid)



But this reply simply fails to tackle the problem. It only suggests that since the problem occurs in other areas of life we should not be overly concerned with it. Thus Lycan basically admits that there is no rationality behind compatibilist accounts of moral responsibility; that compatibilism implies that moral judgements really are arbitrary. I doubt that such an answer would provide much comfort to anyone who raised the objection above.

An answer of a very different sort has been provided by P.F.Strawson. He argues that the question of how we can deserve to be praised or blamed in a deterministic universe is beside the point, since we simply could not give up such 'reactive' attitudes. We cannot help but show resentment, anger, gratitude, blame, and so on. Therefore we do not require any justification for these attitudes.

According to Strawson, if we were to consider abandoning such reactive attitudes, because we came to believe that we were determined and that such attitudes were therefore inappropriate, we would have to accept that we ought to adopt more objective attitudes towards each other. But this would mean radically altering our perception of ourselves:

To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment...to be managed or handled or cured or trained..

..[The objective attitude] cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. (1982, p.66)

He admits that it is not necessarily impossible that we could adopt objective attitudes towards each other, but he states that he is inclined to think that it is practically inconceivable:

The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them, and being involved in interpersonal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question. (ibid, p.68)

Even if we wanted to abandon reactive attitudes, we could not do so. However, Strawson recognises that this answer is not entirely satisfactory. He states that the real question we need to consider is not a question about what we actually do, or why we do it. It is not even a question about what we would in fact do if a certain theoretical conviction gained general acceptance. It is a question about what it would be rational to do if determinism were true, a question about the rational justification of ordinary inter-personal attitudes in general. In answer to this he provides the following reply.

To this I shall reply, first, that such a question could seem real only to one who had utterly failed to grasp the purport of the preceding answer, that fact of our natural human commitment to ordinary inter-personal attitudes. This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework. And I shall reply, second, that if we could imagine what we cannot have, viz. a choice in this matter, then we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of **this** choice. (ibid, p.74)

For Strawson, then, the only justification required for exhibiting reactive attitudes is the fact that we cannot help doing so. And he is clearly unimpressed by calls for the need for a deeper justification. When replying to criticisms of this sort made against his original paper on reactive attitudes he again restated this view:

What I was above all concerned to stress was that our proneness to reactive attitudes is a natural fact, woven into the fabric of our lives, given with the fact of human society as we know it, neither calling for nor permitting a general 'rational' justification. We can see where the limits of our proneness to these attitudes tend to fall, we can understand why they tend to fall where they do and we can find room for the idea of criticism, or appropriateness and inappropriateness, in particular cases. That is all; and that is enough. (1980, p.265)

Strawson's argument rests upon the 'fact' of our proneness to reactive attitudes, which he claims is woven into the fabric of our lives. And this is why his argument fails, for, pace Strawson, there is no such fact. On the contrary, there is substantial evidence to suggest that we can give up reactive attitudes if we believe that they are not justified. The clearest evidence comes from Paul Breer, who specifically argues in his recent book 'The Spontaneous Self' that we have no justification for holding reactive attitudes towards each other and that we should therefore adopt more objective attitudes. He describes in detail how such objective attitudes can be achieved, based upon his own experiences, and explicitly rejects Strawson's arguments.

Breer argues, influenced in part by Buddhism, and in part by science, that we should abandon the traditional notion of agency - that there is some originating 'self'

which alone gives rise to our thoughts - in favour of a naturalistic explanation. What is the outcome of this?

Dispelling the illusion of agency means replacing the notion that we cause our own behaviour with the more realistic idea that all behaviour arises spontaneously out of genetic and environmental circumstance. If it is the circumstances of our birth and training rather than the agent/soul within us that causes our behaviour, we cannot be held morally responsible for who we are. (1989, p.82)

With this last point in mind Breer concludes that we are not justified in exhibiting reactive attitudes and that we should therefore abandon them:

Despite his obvious sympathy for a more objective social order, Strawson seems convinced that it is out of our reach. Since he cites no evidence for his pessimism, it seems likely that he is generalising from personal experience. While I share his observation that most interpersonal relationships are saturated with pride, resentment and other reactive attitudes, I am not convinced that those attitudes are impervious to cultural influence...My own experience over the last eight years has convinced me that reactive attitudes and agency go hand in hand. As our belief in originative self is replaced by a view of experience and behaviour as self-arising, our reactive attitudes give way to more objective ones. (ibid, p.209)

While anger, pride, and despair have continued to arise, I find that they arise less frequently than before and when they do, they rarely last very long. As a result, I feel less buffeted about, more stable, less given to either elation or sadness. (ibid, p.210)

Perhaps the most telling indication of Breer's success in achieving an objective attitude is his remark that he has at times been accused of being a robot, the quintessential objective agent, by his colleagues (ibid, p.211.)

Although Breer represents the clearest example that one could wish for to show that Strawson is wrong in his belief that we cannot abandon reactive attitudes, he is not the only one. Galen Strawson discusses the Buddhist doctrine of 'satkayadrsti', the false doctrine of the

self, whose adherents deny that there is a persisting individual self (Strawson, 1986, pp.117-120). Like Breer, they attempt, through meditation and other means, to achieve an objective attitude to themselves and to those around them. Strawson states:

It is not implausible to suppose that Buddhist monks and other mystics have succeeded in altering quite profoundly their experience of themselves (and others) as acting, thinking, and feeling things.

And - finally - it is not implausible to say that they have...thereby come to adopt the objective attitude. (ibid, pp.119-120)

Clearly, then, there are examples which show not only that it is possible to achieve an objective attitude, but, more importantly perhaps, that it is held to be desirable to achieve such an attitude. Consequently, if we are to solve the problem of whether we can really deserve praise or blame if we are determined, then we cannot do it merely by denying that there is a problem; by denying that we need to justify why we praise and blame people.

We can overcome the problem once we realise that, as morally responsible agents, we choose to act immorally **for various moral and non-moral reasons** and not just because we are determined. The objections in question fail because they do not take account of the role of the agent's reasoning in his acting. It is this reasoning which the attitudes in question are directed at, and this is why such attitudes are justified, even if we are determined.

Consider the objection raised at the beginning of this section as it applies to the feeling of guilt. If you are morally responsible you ought to feel guilty for your immoral actions. This is agreed by both compatibilists and libertarians. But if all our actions are determined, say

by the laws of nature and events in the distant past, then why ought we to feel guilty for our immoral actions? If my robbing the bank was determined, then I could not have done anything else but rob it, so why should I feel guilty for this, when I could not have done anything else? The only conclusion is that if we are determined feeling guilty is unjustified, and a part of what it is to be morally responsible cannot be achieved.

This objection is only superficially plausible. We feel guilty for our immoral actions because we are morally responsible agents. We freely chose to act immorally, and we know that from a moral standpoint we should not have done so, and so, being morally responsible agents, we regret it: we appreciate that we ought not have acted immorally. Determinism does not come into it. Our choice was the result of a variety of reasons, moral and non-moral, but the fact that we were subject to determinism would not have been among them (even if we are determined, we act for reasons, and not just because we are determined.) When we are asked 'Why did you do it?' we will answer in terms of the various moral and non-moral reasons, and not in terms of the fact that we were determined. Guilt is just a natural outcome of our desire to act morally combined with a failure to live up to that desire on certain occasions. It would be irrational of us to try to stop feeling guilty just because we believed that we are determined, when the fact that we are determined plays no part in our choosing to act immorally. Thus even though I may be determined to act immorally on certain occasions, because I freely choose to act

immorally on those occasions, I can rightly feel regret. It was my free choice, as a result of my deliberation, that lead me to act in the way that I did and not the fact that I was determined.

Now consider the related attitudes of praise and blame. We blame certain agents for their immoral actions because they are morally responsible agents. They freely chose to act immorally, and we know that from a moral standpoint they should not have done so, and so, since we know that they are morally responsible agents, we disapprove of them: we appreciate that they ought not have acted immorally. Even if determinism is true it is natural for us, as moral agents, to disapprove of immoral acts by other moral agents because we appreciate, like the agent himself, that he acted for various moral and non-moral reasons and that he ought not to have acted immorally. By blaming an agent we are doing nothing more than finding fault with his moral reasoning. We are not presupposing that determinism does not hold.

It may be objected that such a reply fails because 'ought' implies 'can'; that it only makes sense to say that an agent ought to have acted morally if he could have acted morally. But if he was determined, he could not have done. But this is simply not true, in the sense relevant to ascriptions of freedom, that is, if the agent had epistemic alternatives open to him. When the agent came to make his choice, he could have decided otherwise, for all he knew, and yet in spite of this, in spite of the fact that he could have acted morally, for all he knew, he still chose to act immorally. Even though he was under no

compulsion to act immorally (for no external agent or force was acting upon him so that as far as he knew he could not have decided otherwise) he still decided to act immorally. If we had been there at the time we might have told him that he ought to act morally, because as far as we knew, he could have chosen to act morally. And if we are justified in saying at the time that he makes his choice that he ought to act morally, we are justified in saying after the event that he ought to have acted morally, even if he was determined to act in the way that he acted. Consequently, morally responsible agents clearly deserve to be praise or blamed, even in a deterministic universe.

Objections such as Kane's, therefore, do not hold. We are right to direct our attitudes at agents who act immorally, because they are the ones who chose to act immorally, even when they knew that they ought not to have done. There is no reason to transfer our feelings towards their parents (although we may feel that they share some of the feelings that we direct towards the agent), nor to transform them into bitterness or frustration, because there is no suitable agent to direct them at. It was the agent's own moral reasoning that led to the act being committed, and so it is he who deserves to be blamed for it, even if he is determined.



## Chapter Eight

### The Evolution of Compatibilist Free Will

In this chapter I shall discuss the evolution of compatibilist free will. Unless I specify otherwise free will refers to compatibilist free will in the remainder of this chapter.

I stated in chapter 5 that there are two questions which any account of free will must answer, or attempt to answer. These are:

1) How did free will evolve? That is, by what mechanism or process did free events evolve?

2) Why did free will evolve? That is, what conditions brought about the evolution of free will? What advantages did free will bring to those individuals in which it evolved?

As far as the first question is concerned, the compatibilist is greatly limited in his answer by science's comparatively modest understanding of the brain. At the moment we do not really know in any detail how consciousness, and therefore free will, arises from the electrochemical interactions of neurons in the brain. Consequently, whilst we may be able to draw quite an accurate picture of the structural evolution of the brain, we are limited by the extent to which we can apply this picture to consciousness, and therefore to free will. In a way, the compatibilist is faced with as great a problem as

is the libertarian, who is also unable to explain how (libertarian) free will can have evolved.

However, as far as the second question is concerned, whereas the libertarian cannot explain why libertarian free will should have evolved - why it is advantageous to have the ability to originate one's decisions and the ability to decide otherwise - the compatibilist can explain why compatibilist free will should have evolved.

Basically, in order to explain why free will has evolved the compatibilist needs to explain why it is advantageous for an individual to be aware of the epistemic alternatives open to it, and to be able to rationally evaluate those alternatives. At first glance this may appear to be unproblematic. The individual with the greater ability to rationally evaluate the epistemic alternatives open to it would surely be able to plan its future, and therefore compensate for any problems that may be 'around the corner', more successfully than its rivals (i.e. similar organisms who occupy the same environment, have the same predators, food sources, and so on.) Unlike its rivals it would have the greater ability to appreciate the alternatives open to it; it would have the greater ability to evaluate the best possible course of action out of the various alternatives open to it; it would not be as fixed in its behaviour as its rivals, in that it would be able to respond to problems in a wider variety of ways than its rivals, and hence have a greater chance of finding a successful response; it would, in short, have the greater ability to adapt to changes in its environment. It would therefore be more likely to

survive than it's (less free) rivals. This would explain why free will has evolved. *J. L. Mackenzie*

However, plausible as this explanation is, as it stands it ignores an obvious and important point. This is that for many species the ability to rationally evaluate epistemic alternatives would be quite unnecessary as far as the matter of successfully dealing with future problems was concerned. It would only unnecessarily complicate their lives if they had such higher cognitive powers. In fact, the majority of species do not appear to deliberate over their future actions in the manner that free agents do. Yet the very fact of their existence proves that they are survivors; that they had the ability to out-compete their rivals, and without free will. So why should one regard free will to be advantageous?

Of course free will is necessary to us now, and if we did not have it we would be at a disadvantage, but we now lead highly complicated lives compared to our hunter-gatherer forbears. The compatibilist needs to explain why it was necessary, and therefore advantageous, to them, when it appeared to be unnecessary, and so offer no advantage, to the myriad of other species that occupied a similar environment.

In other words: if free will was so advantageous, then 1) why did it not evolve before?; and 2) why has it not evolved in other species? *make it here* If free will is associated with high intelligence, as it surely must be, and if high intelligence is associated with a well developed brain, which also seems a better than fair assumption, then, given that no animals prior to our ape ancestors had well

developed brains, it seems that no animals prior to this had free will. In other words, the compatibilist must explain why free will has evolved so comparatively recently, and not eons ago, and why it was so advantageous to us, or rather our recent ancestors, and not to any previous species, such as the dinosaurs?

One obvious reply is to state that since free will requires a certain level of intelligence (as we saw in chapter 6) and that since no other animal prior to us, or our recent ancestors, evolved this level of intelligence, that free will simply could not have evolved in any species prior to us. But this just begs the question: why has high intelligence not evolved in any species prior to us, or our recent ancestors? Why did other species not need high intelligence when our ancestors apparently did? This is the crux of the problem, for the answer to this question also provides us with the answers to those concerning free will, as I shall now show.

A number of scientists have recently been struck by the fact that certain laboratory experiments appear to indicate that higher primates are more intelligent than one would expect them to need to be. Chimpanzees, Gorillas and Orang-utans all appear to be able to learn and understand sign language; chimpanzees can (successfully) play computer games; and many higher primates can solve complex puzzles that require the ability to see several moves ahead in a sequence (Gardner, Gardner, & Van Cantfort, 1989; Leakey & Lewin, 1992; Fouts & Fouts, 1993.) As a consequence of this Richard Leakey has stated: The cognitive skills displayed by higher primates in the laboratory seem to outstrip by far the practical demands

of their natural worlds. Has natural selection been profligate in making them smarter than they really need to be? (Leakey and Lewin, 1992, p.284)

Nor is this 'over-intelligence' confined to non-human primates: according to Nicholas Humphrey human hunter-gatherer societies do not use techniques which are greatly superior to those of other social carnivores, and their gathering strategies are comparable to those of chimpanzees and baboons (quoted in Leakey & Lewin, p.285.) But we can also solve complex and abstract mathematical equations and write great symphonies.

So why have all the higher primates, including humans, evolved what appear to be unnecessarily complex intellectual abilities? The answer, says Humphrey, is because of their social lives. All higher primates lead complex social lives and this requires a complex brain. This idea is supported by several recent studies (e.g. de Waal, 1982; Byrne & Whiten, 1988; and Cheney, Seyfarth & Smuts, 1989.) Cheney et al. concluded that 'among human primates, sophisticated cognitive abilities are most evident during social interactions.' 1989, p.1361)

But why, asks Leakey, are primate social lives so complex that they require 'sophisticated cognitive abilities'? He provides the following answer:

In a word, the principle element is **alliances**. As in all animal groups, the ultimate driving factor in individual behaviour is reproductive success. In anthropomorphic terms, females strive to raise to maturity as many offspring as they can; males strive to father as many offspring as they can. For females, reproductive success is achieved through being able to care for and protect offspring; for males reproductive success depends on having as many mating opportunities as possible. For both males and females, the goals are made easier if they can rely on the support of others, friends and relations. A great deal of primate life is therefore spent in nurturing such alliances for oneself and in assessing the alliances of one's rivals. (ibid, p.287)

It is these alliances that require the complex brain. An interaction between just two primates may have considerable repercussions for the other primates in their group:

Hostility between two animals often expands to include whole families, so not only must monkeys predict one another's behaviour, but they must assess one another's relationship...A monkey confronted with all this nonrandom turmoil cannot be content with learning simply who's dominant or subordinate to herself; she must also know who's allied to whom and who's likely to aid an opponent. (ibid, p.289)

Furthermore, the alliances between individuals are always subject to change. The young become adult, adults grow old and die, individuals join or leave the group: any of these changes may provide an individual with the opportunity to form a better set of alliances than it already has.

It is these alliances which provide the answer to the question of why free will first began to evolve in our ancestors, and why it has evolved to such a high degree in humans.

Social primates can only form successful alliances with others if they have a degree of free will. To form successful alliances you need to be able to assess the range of possible alliances that are open to you, and to be able to judge between these alliances in order to determine the best alliance(s) to make. That is, you need to be able to evaluate the possible consequences of making each of the possible alliances open to you. This in turn requires that you believe that you can form any of the possible alliances: it requires epistemic alternatives. Alliance formation therefore specifically requires a

certain degree of free will. And the higher your degree of free will, the better you will be at alliance formation than your rivals, and the greater the advantage you will gain over them.

To form a successful alliance an individual needs to be able to appreciate that it could be advantageous for it to form an alliance with 'A', and that, as far as it knows, it can form such an alliance, and to be able to appreciate that it could be advantageous for it to form an alliance with 'B' instead, and that it could, for all it knows, form such an alliance, and to be able to choose between these alternatives in the most rational way. Successful alliance formation would not be possible without these abilities, without a certain degree of free will. If one could not appreciate that one had a variety of alternatives - different possible alliances - open to one, or if one did not really believe that one could form whichever alliance one chose to form, or if one was incapable of choosing an alliance partner rationally - if one was led by the colour of their eyes, or the size of their feet, rather than by their strength, or one's knowledge of their alliances - then one would simply be incapable of forming successful alliances.

Individuals who do not participate in complex social groups, with their almost constantly shifting alliances, do not need a high degree of free will, and in many cases no degree of freedom at all. They do not have the sort of alternatives open to them that social animals do. The lives of such animals are too constant or repetitive to require them to carry out detailed evaluation of the

future consequences of their actions in the same way that social organisms have to. They can maintain the same pattern of behaviour without threatening their survival. But social primates cannot do this. Their lives are not constant or repetitive and so they do need to evaluate in detail the future consequences of their actions. They cannot maintain the same pattern of behaviour without threatening their survival. In particular, they have to constantly monitor their alliances and make changes when necessary - a friend may become a foe, an inferior may become a superior, and so on. If one needs to be aware of the possible consequences of one's actions virtually all the time, then it is in one's best interests to be as competent as possible in evaluating those consequences. This is provided by free will.

This is why free will has evolved to such a high degree in humans. Those individuals that were best able to determine the alternatives available to them, and to choose the best, would have survived at the expense of those who were less able to - so a high degree of free will would have been clearly advantageous in large social groups composed of alliances, but superfluous to non-social individuals, or social groups with no alliances. By the time primates evolved who were able to communicate through a language, social interactions would have been very complex, and the ability to form advantageous alliances would have been invaluable. This is why humans have the highest degree of free will. Our alliances are the most complex: we have the most alternatives open to us, and they require the most thorough evaluation.



I believe that this is the best explanation for the evolution of compatibilist free will. It not only shows why it evolved, but that it must have evolved, in that we could not have become the agents that we are without it. It should also be clear that libertarian free will adds nothing to this picture. It would not improve our ability to form alliances if we had physical alternatives open to us as opposed to just epistemic alternatives. This is because deliberation - such as the sort that is required to determine which alliances would be the best to make - only requires epistemic alternatives and not physical alternatives (as argued in chapters 2 and 6.) In other words, the libertarian cannot use the explanation I have outlined above in order to explain why libertarian free will evolved, for it explains no such thing.

## Chapter Nine

### Objections and Replies

In chapter 1 I mentioned the fact that both compatibilism and libertarianism have recently come under a great deal of criticism from a number of philosophers who are of the opinion that we do not have free will in either a libertarian or a compatibilist sense. Such criticisms must obviously be countered by the compatibilist and this will be the task of this chapter.

#### 9.i Ted Honderich's Objections

Ted Honderich has recently provided highly detailed critiques of both compatibilism and incompatibilism (1989b, 1993.) His basic objection to these two doctrines can be presented in two fairly concise parts.

In the first part, he argues, by asking us to examine our own feelings in various different situations, that each of us can have two different sets of attitudes, and so make two different sorts of responses, to the likely truth of determinism:

Each of us can focus on either of two conflicting sets of propositions, ideas, or images about actions. One set of these things has to do with voluntariness or willingness - in one of several summary definitions, they have to do with action issuing from embraced desires. We can take these propositions as the only essential ones entering into life-hopes, personal feelings, knowledge, and moral matters. If we do this, we may make the intransigent response to determinism, that it does not matter. On the other hand, we can focus on a larger set of propositions and the like about actions. They have to do with both

voluntariness and origination. We can take it that only all of these considerations together provide good reasons for life-hopes and so on. If we do this, we may make a different response to determinism - dismay. (1988b, pp.118-119)

Honderich stresses that he is not trying to argue that some of us can take one of the attitudes, and others of us can take the other attitude, but that each of us can take both attitudes:

Each of us has, or at the very least is capable of taking, both attitudes, and each of us makes, or is capable of making, both responses. (ibid, p1.3)

I have to say that whilst I agree that each of us can take both sets of attitudes and make both sorts of responses I am not convinced that we ordinarily do. I am not convinced that we are influenced by both sets of attitudes in our day to day thoughts, and so move from one sort of response to the other. As I stated in chapter 1, in my experience the great majority of ordinary people, those who have no knowledge of the free will debate, and so have no philosophical axe to grind, respond to the likely truth of determinism with dismay. Very few indeed appear to be untroubled by it. However, since this is a point which can undoubtedly be debated endlessly, and since it is not of great significance to the general argument, I shall not pursue it. The point to stress, and which I am in agreement with, is that each of us is capable of taking both attitudes and of making both responses.

The second part of Honderich's argument involves a detailed discussion of the ideas of the two main protagonists in the free will debate - compatibilists and libertarians. He discusses the ideas of Hobbes, Hume,

Bramhall, Kant, and several more recent philosophers, and concludes that there are certain propositions which are common to them all:

1. They agree that we all share some single settled idea of what has to be true of a choice if it counts as free, and hence of what has to be true of an action if it counts as free. They say this single idea about the initiation of choices, since we all agree on it, is written into our language.

2. Compatibilists say that our single settled idea of a free choice is of a choice that is according to the desires of the chooser. It is what the chooser really wants. So with a free action. A free choice or action is essentially what was labelled a voluntary one in the last chapter: it is according to the agent's desires and true nature, not against them. Incompatibilists disagree and say that what we all think is a free choice is not only one that the chooser in his true nature really wants but also one that is owed to Reason or the Faculty of the Will or whatever. A free choice is a voluntary and an originated one.

3. Both sides agree in assigning to all of us a certain belief, which they take to be a plain truth. It is the factual belief that something is necessary for something else. A free choice is necessary for holding the person responsible. The sides differ, as just remarked, about what we are supposed to take a free choice to be.

4. Incompatibilists say, as a result, that we all know that people are only morally responsible if determinism is false. Only then can there be choices that are both voluntary and originated. Compatibilists say differently that we all know that people can be perfectly morally responsible even if determinism is true. All we need for responsibility is a voluntary choice.

5. Both sides agree that the question they are concerned with is a logical or an intellectual or theoretical one. What we have to do is just see clearly, not get confused, get a good definition of the idea we all share, not get led astray by other philosophers with a doctrinal axe to grind, check what is or isn't consistent with what, pay attention to this or that proof of what freedom involves. The question is importantly a linguistic one. What we have to do is analyse 'free' in ordinary English and similar words in other ordinary languages. (1993, pp.100-101)

Honderich states that although slight differences do exist between different conceptions of compatibilism and different conceptions of libertarianism 'in the five propositions above we have an accurate summary of the two

traditions.' (1993, p.101) By drawing the two parts of his argument together he is then able to show that the two positions are both mistaken. For despite the assertions of the protagonists, since both libertarianism and compatibilism depend upon attitudes, and not propositional facts, there is no **one** thing in question with respect to what is called our ordinary idea of freedom. Our ordinary idea of freedom involves two sets of attitudes and two sets of responses to the likely truth of determinism. So, he concludes:

If there isn't one thing, then saying that our ordinary idea of freedom either is or is not compatible with determinism may be perfectly pointless and in fact as good as false. (1993, p.101)

When presented in this way compatibilism and incompatibilism are clearly both false. There is no one settled conception of what a free choice is, and so any theory that begins by stating that there is must be wrong. Roy Weatherford has favourably discussed Honderich's argument and echoes his conclusion when he states:

The mistaken belief in a unitary conception is common to both sides (though they disagree on what the unitary conception is) and it is this that makes both positions false. (1990, p.167)

However, it should be clear from the way in which I have approached the problem that one need not define compatibilism and libertarianism so that they both presume that there is a unitary conception of free will. One can define them, indeed, one should define them, as two different interpretations of our ordinary conception of freedom. In which case they are left quite untouched by Honderich's criticism.

In order to illustrate this most effectively I shall

reiterate the general argument of the thesis up to this point. This will also be useful for a further point that needs to be made in a moment. I argued in chapter 2 that we ordinarily distinguish between those people that we take to be free and those we do not. And that this distinction can best be explained by distinguishing between those agents who are in control of their decisions and those agents who are not; between those agents who we take to have alternatives open to them, and to be able to make rational use of them, and those who do not have any alternatives or are unable to make rational use of them. This conclusion is supported by our observations of the behaviour of others and our observations of ourselves; what I have termed our own experience of freedom.

I continued by arguing that we can interpret such observations, including our own experiences of freedom, in two quite different ways, one of which is consistent with our being determined (the compatibilist's position) and one of which is inconsistent with our being determined (the libertarian's position.) Both interpretations have equal right to be regarded as our ordinary notion of free will because both are in accordance with our experiences and because **there is no other acceptable means by which to determine what our ordinary notion of free will is.** In other words, contrary to the majority of philosophers on this subject, and as Honderich rightly argues, there is no one notion of free will that can rightly be called our ordinary notion; there are in fact two such notions. Or rather, there are two interpretations of our ordinary notion of free will.

Libertarian interpretations of our experiences of freedom require that one set of conditions must be satisfied if we are to be free (the U-condition and the C-condition), whilst compatibilist interpretations of our experiences of freedom require that a different set of conditions must be satisfied if we are to be free (the autonomy variables, for example.) Thus the question of which interpretation of our ordinary conception of free will is correct can be settled by a consideration of whether or not we can fulfil these various libertarian and compatibilist conditions. Such a consideration, taking into account various scientific and philosophical arguments, strongly indicates that the conditions of libertarian free will cannot be met, whilst the conditions of compatibilist free will can be met (as I argued in chapters 3 to 8.) Thus the compatibilist interpretation of our ordinary conception of free will should be regarded as the correct one.

This is to my mind the most natural way in which to present the problem, and as such it clearly shows that we do not need to define either compatibilism or libertarianism such that they presuppose that there is only one correct interpretation of free will.

To come to the further point mentioned above, when presented in this way it should also be clear, pace Honderich, that neither libertarianism nor compatibilism depend only upon attitudes or feelings for their support, but rather upon facts. The fact that we take certain people to be free and others not. The fact that we can explain this distinction in terms of self control, of

certain people having alternatives open to them and being able to choose between them in a rational or non-wanton way. The fact that self-control as so defined is open to two different interpretations, which require certain conditions to be met, conditions which can be shown, by various scientific and philosophical arguments, to be either possible or impossible.

It is a fact that we take certain people to be free and others to be unfree. Whether we should or not without any sort of investigation into the justification for this distinction is quite another matter, and we may not always use the terms free and unfree to make this distinction, or even be explicitly aware that we categorise people in this way, but that this is something we clearly do I take to be obvious. To put it another way, we naturally exhibit reactive attitudes towards some people and objective ones towards others (Strawson, 1962.) This is a fact of life if you like. It is also surely a fact that this distinction can be captured by distinguishing between those agents who are in control of their decisions and those agents who are not; between those agents who we take to have alternatives open to them, and to be able to make rational use of them, and those who do not have any alternatives or are unable to make rational use of them. Our observations of others and our own experiences of freedom support this conclusion.

I also take it to be a fact that we can interpret our observations of others and our experiences of freedom in two ways: one which supports libertarianism and one which supports compatibilism. This conclusion is certainly not



owed to a feeling or an attitude. We can explain our experiences of freedom in terms of our having the ability to decide one thing or another there and then (the C-condition), and we can explain these same experiences in terms of our not knowing what the future will hold. This is not like saying that we can explain a doctor's high salary by pointing to the hours he had to work in order to qualify. It is not an explanation in terms of loosely (or even strongly) justified feelings or attitudes. It is saying that we all share the same experience - that we feel that we can make 'there and then' physical choices - and that there are clearly only two explanations for this: either that we can make such choices, or that we cannot, but have this feeling since it is a necessary part of rational deliberation.

Thus it is also a fact that these two different interpretations require certain conditions to be met, conditions which can be shown, by various scientific and philosophical arguments, to be either possible or impossible. In short, it is my contention that attitudes and the like, although they may influence our reasoning in these matters, need not, and really should not. We are, as a matter of contingent fact, able to make free decisions. Our belief in this matter is not a belief without a truth value, it is not owed to a feeling or an attitude, but rather to a set of facts.

One final point. As to the appropriate response for the compatibilist to make to the likely truth of determinism, I do not believe that it is that of intransigence. This is because, as I stated above, I

believe that our ordinary, or prephilosophical, or common-sense attitudes to life are clearly libertarian. Therefore any realisation that we are free in a compatibilist sense cannot lead to the response of intransigence because it will be accompanied by the realisation that we are not free in the libertarian sense that we believed that we were, a realisation which cannot leave our naturally libertarian attitudes untouched. Thus on this point I am in agreement with Honderich. The appropriate response to make to the likely truth of determinism is neither the intransigent response of the traditional compatibilist or the dismayed response of the incompatibilist, but rather 'the response of affirmation.' (1988b, p.148).

This response, in brief, is in three parts. The first part accepts that determinism affects what Honderich calls our life-hopes. (To feel a life-hope is to contemplate one's own future in a general way. It is, amongst one's hopes, 'the dominant one whose realisation is taken at the time as what would make one's life or a coming part of it fulfilled, happy, satisfactory, or anyway of worth.' (1988b, p.14).) Specifically, the response of affirmation causes us to give up those life-hopes that presuppose that we are the originators of our decisions. It therefore contrasts with the response of intransigence.

The second part of the response of affirmation asserts that determinism does not go the other way and destroy life-hopes. It therefore contrasts with the response of dismay. The third part of the response is that those life-hopes that remain untouched by determinism remain life-sustaining things, things of value, since many

of our hopes for the future remain entirely untouched by determinism.

Consequently, my overall conclusion is that although, unlike Honderich, I believe that we are free in a meaningful sense, I also believe that the response of affirmation is generally akin to the sort of response to determinism that I believe is the only one that the compatibilist is justified in making.

### 9.ii Richard Double's Objections

Richard Double has recently provided one of the most original criticisms of free will in his book 'The Non-Reality of Free Will'. He states that his position is quite unlike that of the other four main positions - compatibilism, libertarianism, hard determinism, and the unnamed position which holds free will to be incompatible with both determinism and indeterminism - in that he does not question whether free will exists, but rather he denies that free will is a logically coherent concept:

My non-realist view of free will must be distinguished from the easily conflated position that we are all unfree. My aim is not to subsume one part of the free-unfree dichotomy under the other but rather to undermine the distinction altogether...I do not claim that we are all unfree, but that we are neither free nor unfree in the important sense that the standard positions presuppose. (1991, p.8)

Double states that he does not mean to imply that we do not enjoy many of the capabilities that are typically associated with free will - the abilities to act rationally, voluntarily, non-compulsively, and so on - nor that there are not paradigms of unfree agents such as drug addicts, psychotics, and so on. Rather, he states:

My position is that if we go beyond the clear exemplars of free and unfree agents to ask whether these exemplars stand for distinguishable non-linguistic classes of types of entities, there is ample reason to conclude that there can be no such classes. Not all exemplars that enable us to use terms stand for objective classes of things. (ibid, pp.8-9)

Double's argument for the non-reality of free will is presented in two parts. In the first part he argues that the concept of free will is not a well behaved exemplar concept. An exemplar concept is one that is represented or illustrated by various exemplars. Thus the exemplar concept 'bird' is represented by exemplars such as robin, eagle, chicken, and also, say, 'Fluffy', your pet canary. Free will, argues Double, is not a well-behaved exemplar concept like bird, because free will has multiple, conflicting exemplars. Consequently, a coherent account of our intuitive notion of free will is impossible. He states that ordinarily we do not realise this, but he discusses a series of issues which make explicit the conflicting intuitions which we hold because of these conflicting exemplars. These issues include: (1) whether persons must meet the normative requirements of free will; (2) whether bribes ever reduce our freedom; (3) whether we are made unfree by forces that strongly influence, without absolutely dictating, our choices; and (4) whether being free requires feeling free.

According to Double these four issues illustrate situations in which we can see both that the person could be free and that the person could be unfree, because of the conflicting nature of the exemplars that influence our concept of freedom. Although our inclinations may be swayed one way or the other by the particular exemplars

that influence us, there still appears to be no clear cut answer to the problems the issues raise: there appears to be no 'right answer'. Consequently, argues Double, free will is not a consistent exemplar concept. In which case it appears to follow that free will cannot exist, for it does not seem possible that something can exist, as a real, objective phenomenon or entity, if it has clearly incompatible exemplars.

The second part of Double's argument, which follows on from the first part, is that not only is free will an inconsistent exemplar concept, but it is a subjective, ultimately attitudinal concept as well:

The debate between the compatibilists and incompatibilists can have no resolution because there can be no objective properties of free will or moral responsibility for their accounts to hit or miss. Hence, compatibilism and incompatibilism, contrary to appearances, **cannot** logically conflict, since both positions simply express subjective attitudes. I call this higher level 'compatibility' of compatibilism and incompatibilism **meta-compatibilism**. (ibid, p.133)

#### 9.ii.a Free Will as an Exemplar Concept

According to Double, free will is not a well-behaved exemplar concept because it has multiple, conflicting exemplars. In order to make this explicit Double discusses four issues which give rise to the conflicting intuitions we hold because of these conflicting exemplars. I shall describe each of these issues in turn and then present my reply.

## 1. The Normative problem

Double states that the question of how stringent free will's normative conditions must be poses a serious difficulty for the attempt at a unified account:

Should the compatibilist's account be couched solely in terms of psychological states without commitment to their normative suitability or should the account accommodate the normative aspect? If one says that a free agent's reasons need to be rational only to the agent, then freedom risks being a matter of subjective wilfulness. It would be impossible to reject highly irrational decisions as unfree merely because of their irrationality. This is going to make the free will account too wide. But on the other side of the dilemma, building normative criteria into the free will account entails that much of what prephilosophical common sense takes to be clearly free fails miserably. This result would seem to undermine the compatibilists enterprise, since the philosophical account of free will is supposed to reflect prephilosophical notions of freedom to some significant degree, on pain of the charge of redefining **freedom**. So, either way, the going looks rough. (ibid, p.100)

Double provides several examples to illustrate this problem, of which I shall discuss two. They both show that our intuitions about when to say that someone is free pull in opposite directions, thus preventing a coherent notion of free will. The problem is this: should we provide a strict normative account of free will which would overrule our conflicting intuitions at the risk of discounting some of them and therefore producing an account too far removed from our prephilosophical notion of free will, or should we opt for a more subjective account, and risk providing too wide an account of free will?

The first example involves experiments that demonstrate that experimenters can prompt significant shifts in the attitudes of subjects unbeknown to them. Double takes his example from the work of Nisbett and Ross

(1980) and I provide their own description of the experiment:

..subjects are induced by the experimenter to give a talk or write an essay that is inconsistent with their private beliefs. In "sufficient justification" conditions, subjects are given large monetary incentives for such behaviour and consequently attribute their compliance to the incentive rather than to any corresponding private belief. In "insufficient justification" conditions, by contrast, subjects are paid little or nothing for their counter attitudinal action, and, noting no salient external factors sufficient to account for their actions, and wrongly assuming that those actions must therefore reflect corresponding private beliefs, they change their attitudes so as to bring them in line with their behaviour. These subjects thus commit the fundamental attribution error. Had they correctly identified the situational cause of their behaviour, that is, the subtle social pressures to comply exerted by the experimenter and the experimental context, they would have had no reason to change, or even to reassess, their private beliefs. (p.121)

According to Double, these experiments are important because it is unclear whether the insufficient justification subjects came to change their minds freely or not:

On the one hand, it seems that they clearly manifested free will. What could be a better example of a free choice than one made as a result of self-conscious, elaborated reasons?..

..Nonetheless, it is tempting to argue for the other conclusion. Evidently, getting people to write attitude discrepant essays for small or no rewards may enable us to manipulate their views. The experimenters knew better than the subjects did that an attitude shift would occur. The case in which someone else knows better than you do what you will decide is sometimes held to be a paradigm of lack of freedom. (1991, p.102).

In order to better illustrate the conflict involved, and that it cannot be overcome, say, by arguing that the subjects were partially free, Double suggests the following scenario:

Imagine that Professor Manipulator, a respected psychologist at State U, wishes to have capital punishment made legal in the state. The good professor believes that the best way to contribute to this end is to modify the resistance on the part of as many anti-capital punishment students who pass through State U as possible. To this

end, everyone who takes Psychology 100, required of all first year students, is given an attitude questionnaire, and all students who indicate strong resistance to capital punishment are asked to write essays in favour of capital punishment. (We may assume that the professor adequately disguises what is being done). The result of this systematic use of attitude shift experiments is that after thirty years, a capital punishment law is enacted because it faces very little resistance from the State U-educated people who run the state. (ibid, p103)

The question which Double now asks is the following: if you believed that capital punishment was immoral and deserving of blame upon whom would you attach it? Upon the professor, because he clearly and deliberately manipulated his students, and was fully aware of the attitude shifts he would bring about, whilst they were completely ignorant of his designs. Or upon the students, who were not drugged, coerced, bribed, given misinformation, or otherwise beleaguered? Although the students were ignorant of the fact that they were being induced to change their views, we are all of us often ignorant of the factors behind our changes of attitude and yet we do not consider such changes to be unfree.

Double's conclusion is that there are two clear attitudes we can take to this example. One, that the professor manipulated the students, and that he is the one who deserves to be punished, and two, that the students decided to change their minds after careful consideration of the rational evidence that the professor provided them with, and so are therefore free. We are thus faced with conflicting intuitions over the freedom of the subjects involved.

The second example I wish to consider comes from the recent rationality literature and involves the phenomenon



of bias in favour of existing belief. This phenomenon can best be illustrated using two groups of subjects with strong ideological differences (e.g. on whether capital punishment deters murder.) The two groups are shown two apparently authentic, but really fabricated, pieces of evidence, one of which supports the view that capital punishment deters murder, and the other which supports the view that capital punishment does not deter murder. As Double states, common sense would suggest that the subjects would moderate their opinions after exposure to such mixed evidence. However, in actual fact in experiments carried out by Lord, Ross, and Lepper (1979) the subjects assigned greater weight to the evidence that supported their opinion, and ignored the evidence that opposed it, so that the result was a polarization of opinion.

As before, this example raises conflicting intuitions over whether the subjects are free or not:

We have already suggested why the compatibilist should view our epistemic foibles as destructive of free will: viewed from the outside, the spectacle of our insulating our beliefs from disconfirmation and selecting confirming evidence in a biased way must look, again using Dennett's term, ridiculously sphexish. We appear to be pathetic victims, victimized by our psychological hedonism and laziness. But at the same time, common sense can find reasons to deny this conclusion. After all, where's the compulsion in these cases of epistemic foibles? There is no one to blame for the poor epistemic choices except the subjects themselves...Finally, on the view that free will is co-extensive with the range of decisions that are "up to you," poor epistemic choices seem to be free choices par excellence. (Double, 1991, p.107)

## 2. The Bribe Problem

The issue of bribes raises similar problems to that of the examples above. Double discusses a recent argument of Slote's, who states that bribes may reduce our freedom as much as threats do. Slote supports his claim by pointing out the parity between threats and bribes when viewed from the perspective of utility maximisation. Double states:

..viewed mathematically, bribes and threats can be equally seen as creating alternatives that require the same sort of judgement. In a bribe I must choose between retaining my present utility level or accepting an addition, whereas in a threat I choose between my present utility level or risking a reduction. Viewed from the perspective of utility maximisation, it is difficult to see why bribes categorically cannot reduce freedom if threats do. (ibid, p.110)

However, although Double states that he likes Slote's position he also feels that it is not consistent with the common sense view of bribes (as exemplified, for example, by Double's students.) He states:

The response to Slote I am suggesting is that bribes, unlike threats, **never** reduce free will...A bribe expands your options in the sense that now you have one more alternative open to you that is (by the lights of the bribe-offerer, anyway) at least as desirable as the alternatives that you had before the bribe. In the worst case, where the bribe is unappealing, you are no worse off than you were before the bribe was presented. So, either way you have an expansion of options or at least not a diminution of options and, thus, cannot have your freedom reduced by the other. (ibid, p.111)

Consequently, it seems that we have reasons to believe both that bribes reduce freedom, and that they leave it unaffected, and depending upon our intuitions we may support either conclusion, or see the appeal of both and be undecided.

### 3. Do Causes That Incline without Compelling Reduce Free Will?

The title of this section provides the third problem for free will. Double asks us to imagine cases in which we can resist intrusions that threaten to influence our choices, but only with considerable difficulty. For example, the case of the struggling dieter with a weakness for desserts, who, if offered a dessert, can only resist the urge to have it if he summons up all his 'will power'. Imagine that the dieter can make such an effort only 20% of the time; for the other 80% he yields and has the dessert.

Double states that according to the autonomy variable account of free will (Double's compatibilist account of free will discussed in chapter 6) his freedom is considerably reduced, and adds:

Moreover, a slippery-slope argument yields the same conclusion. We can imagine a continuum of cases where the amount of effort I need to make to resist an inducement is increased to the point that, no matter how hard I try, I cannot decide contrary to the inducement. Because I am unfree when my effort to resist is totally inefficacious, surely I am unfree when the effort required is so great that I can manage it only 1 percent of the time, and so on. Thus, although it is a thorny question of just what degree of inclination without compulsion makes us unfree, there is no problem in principle with saying that sometimes we are made unfree in such cases. (ibid, p.112)

However, Double thinks that there may be a valid objection to this conclusion.

Imagine an objector who argues: "As long as it is true that if you tried hard enough you could have resisted the inducements, you are free. It was up to you whether to expend the requisite effort or not, and if you did not, then you were effectively electing to submit to the inducements. Thus, causes that incline without compelling do not make us unfree; only compulsion does." (ibid)

Again, we are faced with the conclusion that there appear to be two equally compelling answers to the problem of whether causes that incline without compelling reduce free will.

#### 4. Does Being Free Entail Feeling Free?

The fourth problem which Double raises is concerned with whether or not we need to feel free to be free. Again, it seems our intuitions can lead us in two different directions. On the one hand, we may agree with Galen Strawson when he states that an agent must feel free in order to be free:

..if, after an agent performs an action, we discover that it really has no sort of conception or experience of itself as able to choose or act truly responsibly, how can we possibly hold it to be truly responsible for its action? Put yourself in this agent's shoes: you act, when you do, with no sense of yourself as truly responsibly free in your choice or action. Can you call a life spent like that the life of a free agent? That is not what it is to be free. (Strawson, 1986, p.302)

On the other hand, we may be pulled in the opposite direction, as Double himself argues:

First, being free and feeling free seem to be two discrete states. From a psychological perspective, feeling free appears to be a radically different sort of state than the state of being free. Feeling free involves conceptual self-recognition, and, as such, appears linguistic in character. Being free is the way that our choices are, and appears not to be linguistic. If this is so, then, by Hume's law, they must be logically separable...Second, and related to the first point, Strawson's view does not seem to be accurate phenomenologically. I think that I can imagine myself choosing freely one time without having a sense of my free agency. I may evaluate the merits of options A and not A by rehearsing the various utilities that attend each, e.g., 'A has advantage p, whereas not A has the advantage q', without reflecting upon myself. Or, I may choose between the two alternatives without going through any consciously accessible evaluation at all. I may just choose. If I can imagine myself as manifesting free will in such instances, it is difficult to see why I

cannot actually do so in the normal course of things.  
(1991, p.114)

As with the other examples, Double argues that there seems to be no intuitively right answer to the problem of whether being free requires feeling free.

This concludes my discussion of the four issues Double raises which are intended to show that we have conflicting intuitions regarding free will.

According to Double these conflicting intuitions arise because we base our views of free will upon multiple, conflicting exemplars. He discusses three of the many possible exemplars upon which we base our intuitions of free will, and describes the different sorts of intuitions which each exemplar raises (ibid, pp.114-130). Problems of conflict arise for the free will concept, he states, because we do not consistently hold just one of these exemplars, but shift from one exemplar to another, or hold several at the same time (c.f. Strawson, 1986, pp.105-110, and Honderich, 1988b, pp.107-119, who both argue that we can hold both compatibilist and libertarian attitudes at the same time). Double further illustrates our commitment to conflicting exemplars of freedom with various pieces of empirical evidence from the psychological literature (pp.119-124).

The consequence of this is that rather than having a fixed set of intuitions concerning free will (which may or may not conflict with someone else's intuitions, depending upon the sort of exemplars upon which they base their intuitions,) we have conflicting intuitions concerning free will. We base our view of free will upon conflicting

exemplars of freedom. This is why the four issues above raise conflicting intuitions and why we cannot reach a satisfactory solution to the free will problem - there is simply no coherent conception of free will which corresponds to our intuitive notion of free will.

#### 9.ii.b Free Will as an Exemplar Concept - Reply

Double's argument rests upon the assertion that since free will has conflicting exemplars a given account of free will such as the autonomy variable account cannot correspond with anything in the real world. At first this may appear to be a fair assumption, for it seems to be logically impossible that one could have a coherent concept that was made up of inconsistent exemplars. However, such a conclusion would be premature since it is possible to conceive of certain concepts which are taken to describe real entities, and yet which at one time had conflicting exemplars. I wish to describe one such example, and so show that from the fact that a concept has conflicting exemplars it does not necessarily follow that there is nothing in the real world corresponding to that concept.

Consider how an educated person may have regarded the concept of the sun in the seventeenth century. On the one hand, he may have been aware of, and have been inclined towards, Galileo's assertion that the earth revolves around the sun, and this would have formed one of his exemplars. On the other hand, his Christian upbringing, and his own visual experience of the sun's movement, could

have inclined him towards the view that the sun revolves around the earth, which would have formed another of his exemplars. He would therefore have been in the position of having a concept of the sun that was composed of conflicting exemplars. But we would not want to take this as evidence that the sun did not exist in the seventeenth century.

Looking back now we can clearly see that one of these exemplars is based upon faulty intuition - the sun does not revolve around the earth - and someone living at the time might recognise that there is nothing which could ever correspond to the concept of sun as it is given by the two conflicting exemplars, but he would not wish to state that it therefore follows that there is no sun.

What this example therefore illustrates with regard to exemplar concepts is the following:

1) An exemplar concept may be composed of exemplars from a variety of different sources, some of which will be more reliable than others, in that they more accurately reflect how things are in the real world. Where we have conflicting exemplars we should therefore question the justification of those exemplars; we should not automatically regard them all as indispensable. From this it follows that:

2) We require a satisfactory way in which to gauge the validity of conflicting exemplars. Once we have it then the conflict can be resolved without necessarily producing

an artificial or ad hoc concept, or changing in any significant way the original concept formed by the conflicting exemplars. From the fact that a concept C has conflicting exemplars it does not follow that there is not something which very closely corresponds to C, even if C itself is refuted by the fact that it has conflicting exemplars. It may be possible to formulate a revised concept C', which, unlike C, does correspond to something in the real world, which does not have conflicting exemplars, and which, because it will still have much in common with the original concept C, can rightfully bear the name of that original concept.

So, with regard to the first exemplar concept of 'sun' as I described it above, it clearly does not correspond with something in the real world, because two of its exemplars conflict. But once one of these exemplars is rejected (i.e. the exemplar which holds that the sun revolves around the earth) then the other exemplars (that the earth revolves around the sun, that the sun is hot, yellow, spherical, etc) remain consistent with each other and allow for a concept that appears to exist in the real world, and which is so close to the first concept that it still justifies the term 'sun'. In this case the exemplar was eliminated because there were good scientific reasons for doubting its validity (after Galileo).

The question is: is this example analogous to the situation as it stands with free will? Can we find a way to judge between the conflicting exemplars of free will so



that we can remove the conflict and still retain an acceptable concept of free will? I believe that we can.

The exemplars for free will can come from a variety of different sources. Double himself states that our selection of exemplars (or paradigms, as he calls them) is governed by both dispositional and situational factors:

By the former I include things like this. A philosopher might be more likely to rely on a certain paradigm more often than another (or to explicitly reject a certain paradigm) because of chronic traits such as temperament, upbringing, ideological bias, or related philosophical views...Also under the heading of dispositional factors is the case of having received reinforcement for or against using one of the paradigms. It is easy to imagine how someone who achieves success in publishing books and articles that exploit one of the paradigms would be reinforced to continue to rely on that paradigm. Having expressed this much sensitivity to the mundane motivations that might govern our paradigm selection, I hasten to add that a philosopher might grow enamoured of a specific paradigm because that paradigm appears more fruitful and continues to withstand rigorous scrutiny.

Under the heading of situational factors, I include a wide range of logically relevant and fortuitous factors that might contribute to reliance on a certain paradigm. Having just read a certain novel (or philosophical work), seen a movie, or having tried to console a depressed friend might make one more likely to opt for a particular paradigm. (ibid, pp.131-132)

Double appears to regard all factors which lead to our choice of exemplars as equally acceptable, and indeed he has to, for if he did not then his basic argument against free will would falter. He argues that a satisfactory account of free will must satisfy all of our intuitions concerning free will, since our intuitions (and the exemplars from which they arise) are the only means we have for determining what free will is (ibid, p.131). However, it seems obvious to me that some of these exemplars will be more reliable or acceptable than others, in that they will more accurately reflect how things are in the real world. After all, we have already seen that

one can have exemplars for the concept 'sun' which do not mirror how things are in the real world. Consequently, if one does not regard all of these intuitions, and the exemplars that give rise to them, as equally acceptable, then there seems to be no reason to require that a concept of free will satisfies them all; it would only need to satisfy the acceptable ones, and these might not conflict, enabling one to reach a satisfactory conception of free will, and so settle the conflicts caused by the 'conflicting intuition' cases described earlier.

But do we have good reason to believe that the exemplars which we use to formulate the concept of free will are not all equally acceptable? I believe that we do, as I have argued throughout the thesis. Those exemplars which find their support in beliefs or propositions which clearly conflict with the current scientific world view are not as acceptable as those exemplars which do not. This enables us to reject certain exemplars of freedom, and so produce a coherent account of free will, without changing in any significant way our ordinary concept of free will (as discussed in chapter 2.) We can then apply this coherent free will account to the various 'conflicting intuition' cases Double describes and achieve consistent and satisfactory solutions.

Those exemplars which make use of the C-condition and the U-condition, for example, or which require a dualist metaphysics, should be rejected, since they are inconsistent with the current scientific world view, as I have argued earlier in the thesis. Those exemplars which do not rely upon such conditions, but upon conditions

which are consistent with the current scientific world view, should be the ones upon which we base our conception of free will, for they are the ones which correspond with the state of affairs in the real world. In other words, the only acceptable concept of free will is one which is formed by the exemplars which lead one to form a scientifically compatible conception of freedom like the one expounded in this thesis: that an agent is capable of free decisions if he is capable of rationally (or non-wantonly) choosing between the epistemic alternatives open to him. This requires, as discussed in chapter 6, that he fulfils the autonomy variables. If we apply this account of free will to the various 'conflicting intuition' cases we can resolve the problems they raise in a completely satisfactory way.

Consider first of all the problems raised by what Double calls the 'normative problem'. According to Double we have two choices: we can either hold that a free agent's reasons need to be rational only to the agent, in which case freedom risks being a matter of subjective wilfulness. Or we can try to build certain normative criteria into our free will account, in which case much of what common sense takes to be clearly free fails miserably. Either way, says Double, the going looks rough.

But once we accept that the exemplars from which our intuitions arise are not all equally acceptable, then we can see that the going is not nearly as rough as Double imagines. We should not view it as a failure if our free will account causes us to abandon some of our intuitions, but rather as a step towards a more satisfactory account.

In fact, we should view it as a success if our account causes us to abandon those intuitions which arise from inconsistent or otherwise unacceptable exemplars.

With this in mind, we can provide perfectly adequate solutions to the two cases which I discuss earlier. As far as the first example is concerned, in which students have their views altered by Professor Manipulator, the students are not free since they did not fulfil the self-knowledge variable. They did not correctly identify the situational cause of their behaviour, that is, the subtle social pressures to comply exerted by the experimenter (Professor Manipulator) and the experimental context. If they had done, if they had been fully aware of all the causes behind their decision, if they had had adequate self-knowledge, they would have had no reason to change their beliefs. Consequently, as to Double's question of who should be punished for the students views, the answer must be the professor. Rather oddly, Double does not consider this response, even though it is the obvious one for the autonomy variable compatibilist to make. On the one hand, when discussing the example of Professor Manipulator, he states that we are all often ignorant of the factors behind our changes of attitude and yet do not consider such changes to be unfree, whilst on the other hand he expounds an account of free will whose self-knowledge variable explicitly holds ignorance of this sort to be incompatible with free will. Given this, it seems curious that he even considers the example to be a problem for the compatibilist at all.

As far as the second example is concerned, the case

involving bias in favour of existing belief, the subjects in question are not free because they were not rational in their deliberation. They completely disregarded the evidence that counted against their beliefs, whilst accepting the evidence that supported them. Double argued that the subjects could be considered to be free: 'After all, where's the compulsion in these cases of epistemic foibles?' (ibid, p.107) But a voluntary choice is not necessarily a free choice, as I argued in chapter 6, and so this is no reason whatsoever to assume that such subjects could be free.

The second general problem case that Double discussed involves bribes. Do bribes increase one's freedom or reduce it? According to Slote, they may reduce it, but according to Double, they could increase it. From the point of view of the account of free will I have expounded, bribes will not affect one's freedom if they do not prevent one from rationally choosing between the epistemic alternatives available to one. Therefore, as far as most, if not all, bribes are concerned, the agent's freedom would be unaffected.

That notwithstanding, I have to say that I do not regard this to be a very sound example of a situation in which we can have conflicting intuitions. It seems to me that Slote's argument that bribes can reduce one's freedom is very weak. Double is surely right when he says in opposition to it that bribes clearly expand one's opportunities whereas only threats reduce them? Bribes do not represent a danger to one's life, career, loved ones, and so on, whereas threats clearly can. Bribes do not

therefore force one down a certain path in the way that threats can. They may be as compelling as threats in certain situations, but such compulsion will arise from the person's character, from his greed or desperation, rather than from some external threat to his or another's life.

The third case involves the problem of whether causes that incline without compelling reduce free will. I think that this case is a special case of a more general problem that can occur in the free will debate: the problem of whether free will is a matter of degree or an all-or-nothing capability. Those who believe free will to be a matter of degree will answer yes to the problem in question - that causes that incline without compelling can reduce our freedom - whereas those who believe it to be an all-or-nothing capability will answer no - that only compulsion can remove our freedom.

We have already seen that for free will to have evolved it must have done so gradually; that it must have evolved by degree. This would seem to point to the conclusion that free will is not an all-or-nothing capability. Furthermore, it seems clear to me that self-control should be regarded as something that admits of degrees. At one extreme we have the individual who is totally under the control of another, such as a hypnotised person. At the other extreme we have someone who is not under the control of anyone at all. But there are a whole range of circumstances in between these two extremes. We may decide to do something because someone has suggested to us that we do it. In this case we still seem to be very

much in control of our decision. Then there are cases where we make a certain decision because we have been persuaded that it is right to do so. Here, although we may still be said to be making a free decision, we may not have made the decision had it not been for the person's persuasive arguments. To say that the person controlled our decision is too strong, but they certainly strongly influenced it. A further situation is one in which we are pressured into making a decision. Such pressures may be light, in which case they do no more than influence our decisions in the manner of persuasion, or they may be quite strong, in which case they come close to controlling our decisions. Indeed, very strong pressures, or threats, can be seen in many cases to amount to control of our decisions.

There is a further complexity to this situation. This is that people vary in their susceptibility to the various influential factors considered above. Weaker willed people will be more strongly affected by persuasion than strong willed people, for example. People with a particular aversion to something will be more strongly affected than people with a lesser aversion, or no aversion at all. What may be sufficient to control one person's decision, to rob them of their freedom, may be insufficient to control another person's decision. All this points to the conclusion that self-control, and therefore freedom, is something that exists in degrees.

However, at least one philosopher has argued that free will is an all or nothing capability (Thorp, 1980, pp.7-9, 137-139.) It can be argued that:

the test for freedom in a decision is 'could he have decided otherwise?', and, whatever that question means, it seems to be one which admits only yes or no as an answer. (p.8)

But this line of argument is surely mistaken. For the ability to make free choices is not simply a matter of whether we have alternatives open to us, but how we make use of them. We may have alternatives open to us, but be so lacking in the necessary abilities to choose between them rationally, that we do not have free will. If we only measure free will in terms of whether or not we could have decided otherwise then we completely ignore this rationality component. Once we allow for rationality, then given that rationality can occur in degrees, we can see that free will must do so as well.

Therefore in answer to the problem of whether or not causes that incline without compelling reduce free will, we can state that they do reduce free will. Arguments to the contrary are based upon the mistaken belief that free will is an all or nothing capability.

Double's last 'conflicting intuition' case concerns the problem of whether or not being free entails feeling free. He states that we can conceive of reasons why we should regard freedom to require the belief that we are free, and reasons why we should not regard freedom to require this belief.

In order to determine whether one needs to feel free in order to fulfil the account of free will expounded in the thesis, we need to determine what it means to feel free. Amongst other things, to feel free is to feel that one has alternatives open to one. It is to feel that one



can freely choose one thing or another. If one does not have this feeling, if one does not feel as if one has alternatives open to one, then one will make no attempt to evaluate one's decisions, since evaluation is unnecessary if one lacks any alternatives to the decisions one makes, and so one's decisions will simply not be free. Thus one at least needs to feel that one has alternatives open to one in order to make free decisions.

Another part of feeling free is feeling that one can decide between the alternatives open to one in an effective way. One feels that it is owing to one's own efforts that one chooses this or that, and not owing to some other factor. If one does not feel that one's efforts are effective in enabling one to decide between alternatives, if one does not really feel in control of one's choices, then one will begin to doubt the point of making any effort, of evaluating the reasons for and against certain choices, and so one's decisions will become wanton, and one's freedom will be compromised. So one needs to feel that one is in control of one's choices to be in control of them.

Consequently, it seems to me that there are at least two good reasons for believing that we need to feel free in order to be free. In fact, pace Double, I find it difficult to conceive of an agent that could be free and yet not feel free. How could an agent make a rational choice from amongst the alternatives available to it if did not feel that it was capable of making such a choice, either because it did not feel in control of it's decision making processes, or because it did not feel it had any

alternatives?

If there is a problem in connection with this issue, it is that the conclusion above conflicts with what Galen Strawson terms 'The Principle of Independence'. This states that believing something to be true is not a condition of that thing's being true (Strawson, 1986, p.15.) This does appear to be a sound principle. It is surely true, for example, that believing that the 'Dog and Duck' is in the high street is not a condition of it actually being in the high street. The belief and the actuality are not connected. But if this is the case, then the assertion that one must feel that one is a free agent is a necessary condition of being a free agent must be abandoned, since it conflicts with the Principle of Independence.

The only way to resolve this problem without abandoning the idea that one needs to feel free in order to be free is to argue that the Principle of Independence is not universally valid. It does not apply to those cases such as free will where the belief that something is true is a necessary condition of that thing's being true. Non-mental entities or phenomena such as the 'Dog and Duck' exist regardless of whether there are any beliefs associated with them (given, pace Berkely, that physical objects exist apart from our ideas about them.) But mental phenomena, such as free will, cannot exist apart from our beliefs about them, and so the Principle of Independence simply cannot apply to them.

This concludes my discussion of Double's first objection to free will.

### 9.ii.c Meta-compatibilism

Double's second main argument against free will involves what he terms the argument for meta-compatibilism. He argues not only that free will is a badly behaved exemplar concept, but that it is a subjective, ultimately attitudinal one as well:

The debate between the compatibilists and incompatibilists can have no resolution because there can be no objective properties of free will or moral responsibility for their accounts to hit or miss. Hence, compatibilism and incompatibilism, contrary to appearances, cannot logically conflict, since both positions simply express subjective attitudes. I call this higher level 'compatibility' of compatibilism and incompatibilism **meta-compatibilism**. (1991, p.133)

Double states that the central aim of his book is to reject the following thesis:

(F) It is logically possible that agents manifest free will in the deep philosophical sense that is believed to warrant moral responsibility. (ibid, p.134)

The argument for Meta-Compatibilism enables this to be achieved, he argues, because it enables one to reject the following thesis, which is equivalent to (F):

(A) There is an answer to the dispute between the compatibilists and the incompatibilists over the correct analysis of free will. (ibid)

The argument for meta-compatibilism can be put in five parts:

1) If there is an answer to the dispute between the compatibilists and the incompatibilists over the correct analysis of free will, then free will (and moral responsibility) are objective properties that persons might instantiate under certain conditions.

2) Free will and moral responsibility, if they exist, are moral properties.

3) If there is an answer to the dispute between the compatibilists and incompatibilists over the correct analysis of free will, then at least two classes of moral properties are objective properties that might be instantiated under certain conditions.

4) There can be no moral properties.

5) Therefore, there is no answer to the dispute between the compatibilists and incompatibilists over the correct analysis of free will. (ibid, pp.134-139)

Consequently, since free will and moral responsibility are subjective properties, the two positions - compatibilism and incompatibilism - cannot logically conflict: they are (meta-) compatible (Double's position thus has certain similarities to Honderich's.)

As Double states, his argument rests upon steps (2) and (4): that free will and moral responsibility are moral properties, and that there can be no objective moral properties. It is step (2) which I shall refute. I shall say nothing about (4). First, because rejecting (2) makes further criticism unnecessary, and secondly, and more importantly, because I think that Double could be right on point (4) (although see Kapitan, 1994.)

Double's argument that free will and moral responsibility are moral properties - step (2) - is given in two stages. First, he argues that the way in which we define moral responsibility clearly shows that it has an undeniable moral character. Secondly, he argues that since free will is so closely associated with moral responsibility, it must also be a moral property. It is this second inference which I object to.

That moral responsibility is a moral property almost goes without saying; we are talking, after all, about moral responsibility. But I do not believe that it therefore follows that free will is a moral property. Here is Double's argument:

Given that we have seen that moral responsibility is a moral notion, and that its application is generally

believed to depend upon whether agents are 'really' free, this provides some reason to think that free is also a moral concept. This is not to claim that all properties entailed by moral ascriptions are moral properties...But with moral responsibility and free will, the dependency is almost mutual: "S did a freely" strongly supports "S is morally responsible for a." It seems that any notion that stands in such an intimate relation to a moral notion will itself be a moral notion.

..A final consideration would be to point out that not only does the applicability of moral responsibility depend upon the applicability of free, but that the same sort of considerations are appropriate in deciding whether to apply either term...In many discussions, considerations regarding responsibility and freedom, although analytically distinguishable, are interchangeable as we appeal to the same factors in speaking to both issues. (1991, p.137)

If you do not look too closely into this argument then it does have an air of plausibility. After all, free will and moral responsibility are often interchangeable, and so it seems acceptable to conclude that if moral responsibility is a moral property that free will must be as well. Upon closer inspection, however, the argument is less convincing.

The reason why free will and moral responsibility terms are often interchangeable is because free will is one of the conditions of moral responsibility. If you did not decide to A freely then you cannot be held morally responsible for A. This is why they have such an intimate relationship. But simply because they have this much in common it does not follow that they have anything more in common. It does not follow from the fact that moral responsibility is a moral property that free will is a moral property as well. Consider what it is about moral responsibility that makes it a moral property. It is because it has a definite moral condition: one requires a sense of morality - of knowing right from wrong - in order

to be morally responsible. But, pace Wolf (1990), to make a free decision you do not need to be able to discern what is morally right from what is morally wrong. Free will has no moral condition. This shows why moral responsibility is a moral property and free will is not.

I therefore conclude that Double's argument for meta-compatibilism does not succeed. He has failed to show that free will is not an objective, non-attitudinal concept. Consequently, I see no reason not to conclude that it is possible to reach an answer to the dispute between the compatibilists and the incompatibilists over the correct analysis of free will.

### 9.iii Bruce Waller's Objections

A more traditional objection to moral responsibility than those so far considered in this chapter has recently been discussed in a number of articles by Bruce Waller (1989a, 1989b, 1993), who supports the relatively uncommon position of hard determinism. His basic objection to moral responsibility can be summed up by the following statement:

Since people start unequally - at starting points not of their own choosing or making - they are not morally responsible (do not justly deserve blame or credit) for their finish (Waller, 1989b, p.209.)

I take this objection to be really two objections. The first states that people should not be held morally responsible because they have different 'starting points'. This is not just to state that people are born differently - that they differ genetically - but also that they are

brought up in different ways; they have different experiences or life histories. They will therefore all have different abilities and opportunities, possibly widely different, and so cannot be held morally responsible:

We are not equal in opportunities and capacities, and moral responsibility judgements based on the assumption that we are roughly equal are both unfair and implausible. (Waller, 1993, p.49)

The second objection states that we should not be held morally responsible because, not only do we have different abilities and opportunities, but we are not ultimately responsible for these differences (in other words, we do not fulfil the U-condition of moral responsibility.) Since I have already considered the problems raised by the U-condition in chapter 3 I shall not discuss it further. Rather, I shall discuss the first part of Waller's objection, which arises from the fact that we have different opportunities and abilities.

I do not disagree with Waller when he states, for example:

Even excluding the 'retarded or psychopathic', we remain vastly different in intelligence, education, fortitude, imagination, inquisitiveness, sympathy... (1989b, p.49)

That people have different opportunities and abilities, and that these differences may profoundly affect the way that different people should be regarded with respect to moral responsibility, I take to be obvious. I basically agree with Waller's criticisms of those, like Dennett (1984) and Frankfurt (1975), who argue that our different opportunities and abilities make no significant difference to our degree of moral

responsibility. That such differences do have a significant affect is often painfully obvious (see Waller, 1989b and 1993, p.49; Klein, 1991; and especially Watson, 1987b, for discussions leading to this conclusion.) But I fail to see why these differences should cause us to conclude that no one should be held morally responsible. They do not give us good reason to reject the concept of moral responsibility.

In chapter 7 I stated that if we make our decisions freely, and if we have a proper understanding of the moral consequences of our decisions, and the actions which will result from then, that we are morally responsible. Given this, we cannot be morally responsible for our actions if 1) they are not the result of our free decisions (e.g. if we had no alternative open to us, or if we were incapable of rational thought), or 2) we do not properly understand the moral consequences of our free decisions; that is, if we do not have a proper understanding of right and wrong. There are no other reasons to deny moral responsibility.

Consequently, although we may have different opportunities and abilities, if these differences do not affect our ability to make free decisions, or our ability to appreciate the moral consequences of our acts, then they will have no bearing upon whether or not we are morally responsible. And there is no reason to suppose that such differences will automatically affect either of these two factors. We have no good reason to believe that people from Africa or Australia, for example - people of widely differing opportunities and abilities - are not able to appreciate the moral consequences of their



actions, to appreciate that they have alternatives open to them, and to choose the best one, in the same manner that we can. So it does not follow that simply because we have different opportunities and abilities we cannot be considered morally responsible. Such differences may affect the degree to which a person is morally responsible, in which case we should make allowances for such people's behaviour, but they certainly need not, and so if one simply presumes that they will, as Waller does, then one is making as grave an error as those who presume that they will not affect them.

Consequently, pace Waller, differences in abilities and opportunities do not provide us with a reason to reject the concept of moral responsibility. However, to state that they do not provide us with a reason to reject the concept of moral responsibility is not to state that they do not affect moral responsibility. As I have already said, I do not agree with those such as Dennett and Frankfurt who have argued that differences in opportunity and ability make no significant difference to our moral responsibility. Whilst these differences do not cause us to reject the concept of moral responsibility they do clearly affect the **degree** to which different people can be held morally responsible for similar acts.

This is because such differences can clearly affect 1) one's ability to appreciate the moral consequences of one's actions; 2) the sort of alternatives that are open to one; and 3) one's ability to appreciate what sort of alternatives are open to one. For example, if one lacked the opportunity of an adequate moral education then one

would not be able to appreciate the moral consequences of one's actions in the way that a morally educated person is able to. Thus one should not be held to be as morally responsible as that person for the same type of acts. If one had a deprived childhood - say, deprived of love, support, or a certain level of education - then one could lack the alternatives that are available to a person who had all of these things in childhood. As a consequence, one may be unable, for example, to form satisfactory relationships with people, or to motivate oneself, or to compete for jobs. Thus one may find oneself, through no fault of one's own, having to resort to certain morally irresponsible acts in order to achieve what others achieve though morally responsible ones. Or one may simply not be able to appreciate all the alternatives that are available to one, and so only choose morally irresponsible alternatives, because they are the only alternatives that one feels are open to one.

It may be raised at this point that this argument represents the thin end of a rather objectionable wedge. That if one holds one person less morally responsible for a certain type of act than another person, that we are allowing in the 'specter of creeping exculpation'. (Dennett, 1984, p.156) This is a very real problem. If we wish to hold certain people less morally responsible than others, then who, and for what? We do not want to give people the opportunity to abuse the fact that we will be more lenient with them for their morally objectionable acts because they lack, or lacked, the opportunities that are open to others. Yet at the same time, we must

recognise that this background profoundly affects the way that they act, and so to hold such people equally morally responsible with those of a different background must be unjust.

Martha Klein suggests that we should adopt what she terms the payment-in-advance principle:

..an offender who has suffered as part of the process of becoming the sort of person who fulfils the M-condition [a person who commits morally objectionable acts] has already paid something in advance for his offence and in virtue of this deserves to pay less for his wrongdoing than someone who has not suffered (or suffered less) but has committed the same sort of offence. (1990, p.160)

However, although such a scheme may be attractive in principle, in practice it faces the problem of how to determine whether someone has suffered sufficiently to deserve leniency for their wrongdoing. What amount of suffering equates with what amount of leniency? Presumably, on this principle, someone who has been regularly beaten as a child deserves leniency when charged with striking out at their own child in a sudden rage, compared with, for example, someone who has not been beaten as a child, and who calculatingly strikes their child not in rage, but as a punishment for some misdemeanour. But what about other wrongdoings committed by the first parent? How much can be laid at the door of the deprived childhood, and how much at the person himself? If he commits a murder, could it really be attributed to his childhood? It seems reasonable to doubt it. Yet the case of suddenly striking the child seems entirely consistent with his childhood history, and so leniency seems appropriate (as well as, presumably, some form of psychological counselling.) But how can we

determine whether the suffering is relevant to the wrongdoing, and how can we prevent abuse of this principle? How can we prevent someone from a deprived childhood of committing acts for which he knows that he will receive leniency?

It has to be said, of course, that merely by being aware of the possibility of abuse of the principle, we immediately reduce the chances of that abuse. Through our experience, through trial and error basically, we can come to know when to be lenient and when not to be. We can learn to minimise the chances of abuse. But I suggest that this is hardly a very satisfactory solution. For we may get it wrong many times before we get it right, as real life cases have all too often shown. It may even be argued that the possibility of leniency encourages abuse of the system, and that it is better to punish, possibly unjustly, and so prevent others from committing similar acts, than be lenient and so possibly encourage others.

For these reasons I believe that the payment in advance principle, and other systems of leniency, do not provide anything more than a temporary solution to the problem. There seems to me to be only one permanent solution. This is to remove, or to reduce as far as is possible, the relevant (relevant to moral responsibility) differences in opportunity and ability that exist between people. If a society wishes to hold people (or the great majority of people) equally morally responsible, which is surely the aim of any society which wishes its members to abide equally by one set of laws, then that society must ensure, as far as is possible, that the people who make up

that society have the same opportunities open to them, and that any relevant differences in opportunity must, as far as is possible, be eliminated. This eliminates, or at least severely reduces, the possibility of certain people being held to be less morally responsible for their acts than other people on the grounds of differences in opportunity or ability, and so prevents them from being able to use their background to excuse their behaviour.

Practically speaking, it may not be possible to remove all the differences in opportunity that exist between people, or even to reduce them to a satisfactory level, and there will always remain certain differences in ability between people of different levels of intelligence. However, there are a great many differences in opportunity and ability - due to differences in education, to name one obvious example - that can be resolved more easily than others. It should therefore be possible to reduce or remove at least some of the differences that presently exist, and this would go a long way towards reducing the general problem.

This concludes my discussion of objections to free will and moral responsibility.

## Chapter Ten

### Conclusions

I stated in chapter 1 that there are three main issues that need to be tackled in the free will debate. Firstly, it must be determined whether or not our personal experiences of freedom point us towards any particular conception of free will; secondly, given that they at least support libertarian free will it must be determined whether or not the conditions necessary for libertarian free will can actually be fulfilled; and thirdly, if they cannot be fulfilled, and if our personal experiences of freedom also support compatibilism, it must be determined whether or not a compatibilist conception of free will can be expounded which overcomes the various objections raised above.

I have discussed these three issues and concluded that our experiences can point us towards either libertarianism or compatibilist, and that whereas we can fulfil the conditions of compatibilist free will, we cannot fulfil the conditions of libertarian free will. Indeed, regardless of anything else which the thesis may have achieved, I believe that it has shown conclusively that libertarian free will is simply not possible.

Consider first the scientific evidence. It remains unclear how libertarian free will can arise from the neuronal processes of the brain, and therefore how it can

have evolved. It also remains unclear why it should have evolved: it appears to offer no advantage to individuals that possess it over those that do not. Finally, and perhaps most damagingly, it is clearly inconsistent with the physicist's nature of time. Scientific considerations apart, the arguments of Frankfurt and Strawson, as well as those I raised in chapter 3, leave little hope that a satisfactory account of libertarian free will can be achieved.

At least, this is the way that I see it. Libertarians will no doubt disagree, but if so then there are a number of problems to which they must provide answers if they wish their ideas to be taken seriously. They must explain how free will can have evolved, and why it has evolved. They must explain how free will arises from the neuronal processes of the brain, and, if it does not, how in fact it does arise. They must show how free will can be a matter of degree, since this accords with our experiences, and with the evolution of free will. And they must overcome the problems raised against the C-condition by Relativity theory and Frankfurt's argument.

In view of the probable failure of libertarian free will to provide a satisfactory account of free will, I believe that a compatibilist account of free will provides the best hope of an answer to the problem of free will. My own view is that an epistemic account based upon Double's autonomy variable account is the most satisfactory, in that it is consistent with our experiences of freedom and with the current scientific world view, and does not suffer from the problems which afflict other compatibilist

accounts of free will. However, whereas I am quite sure that libertarian free will provides no solution to the problem of free will, I am less sure that the compatibilist account expounded provides the solution. I have tried to anticipate objections to such an account, but since it is always easier to see the faults in others' arguments rather than in one's own, it is entirely possible that I have missed something. Having said that, I do believe that the answer to the problem of free will is to be found along compatibilist lines, and so I am confident that I have at least provided a stepping stone towards the solution. And after several hundred years of often fruitless debate, to hope for something more would be somewhat rash.

If there is one weakness in any compatibilist account of free will that needs to be considered above all others it is that compatibilist accounts are not perceived to comply with our experiences of freedom. Despite my arguments to the contrary in chapter 2, I accept that many people will find the idea that compatibilism is consistent with our experiences unpersuasive. I know from experience that many people view the arguments of chapter 2 with suspicion, and for that reason find the idea of compatibilist free will quite unacceptable. The only possible response, beyond restating the arguments in a different form, is to ask the objector why he believes that compatibilist accounts of free will are inconsistent with our experiences. Exactly which experiences is he referring to? Is he guilty of the same misunderstanding which caused people to believe that the sun revolved



around the earth? That is, does he believe that there is only one explanation for our experiences? Are there any sound reasons underlying his objection, or is it rather due to the sort of unsound fears and misconceptions concerning determinism which Daniel Dennett recently discussed? (Dennett, 1984, pp.5-17) By considering these points in detail the libertarian will then be in a position either to state exactly why he believes that he is right to object to compatibilist accounts of free will, or to appreciate that his objections are unfounded. Either way, progress will be made, and one cannot ask for more than that.

## Bibliography

- Ayer, A.J., 1954, 'Freedom and Necessity', in **Philosophical Essays** (London: Macmillan.)
- Adair, R.K., 1993, **The Great Design - Particles, Fields, and Creation** (Oxford: Oxford University Press.)
- Barrow, J., 1988, **The World Within the World** (Oxford: Oxford University Press.)
- Berlin, I., 1954, **Historical Inevitability** (Oxford: OUP.)
- Blumenfeld, D., 1971, 'The Principle of Alternate Possibilities', *Journal of Philosophy*, 68, pp.339-345
- Blumenfeld, D., 1988, 'Freedom and Mind Control', *American Philosophical Quarterly*, 25, pp.215-227.
- Breer, P., 1989, **The Spontaneous Self** (Cambridge, MA: Institute for Naturalistic Philosophy.)
- Byrne, R.W. & Whiten, A., eds., 1988, **Machiavellian Intelligence** (Oxford: Clarendon Press.)
- Campbell, C.A., 1957, **On Selfhood and Godhood** (London: George Allen & Unwin Ltd.)
- Cavalieri, P. & Singer, P., eds., 1993, **The Great Ape Project** (London: Fourth Estate.)
- Cheney, D.L., Seyfarth, R.M., & Smuts, B.B., 1986, 'Social relationships & Social Cognition in Non-human Primates', *Science*, 234, pp.1361-1366.
- Chisholm, R.M., 1966, 'Freedom and Action', in Lehrer.
- Chisholm, R.M., 1982, 'Human Freedom and the Self', in Watson, 1982.
- Churchland, P.M., 1984, **Matter and Consciousness** (Cambridge, MA: The MIT Press.)
- Churchland, P.S., 1986, **Neurophilosophy: Towards a Unified Theory of the Mind/Brain** (Cambridge, MA: The MIT Press.)
- Clarke, R., 1991, 'Free Will and the Conditions of Moral Responsibility', *Philosophical Studies*, 66, pp.53-72.
- Clarke, R., 1993, 'Toward A Credible Agent-Causal Account of Free Will', *Nous*, 27, pp.191-203.
- Crane, T. & Mellor, D.H., 1990, 'There is No Question of Physicalism', *Mind*, 99, pp.185-206.
- Davidson, D., 1979 (1970), 'Mental Events', in Honderich & Burnyeat.

- Davies, P.C.W., 1983, **God and the New Physics** (London: J.M. Dent & Sons Ltd.)
- Davies, P.C.W., 1986, **The Forces of Nature** (Cambridge, Cambridge University Press.)
- Dawkins, R., 1986, **The Blind Watchmaker** (London: Penguin Books Ltd.)
- Dennett, D.C., 1984, **Elbow Room** (Oxford: Clarendon Press.)
- Denyer, N., 1981, **Time, Action and Necessity: A Proof of Free Will** (London: Duckworth.)
- Double, R., 1989, 'Puppeteers, Hypnotists, and Neurosurgeons', *Philosophical Studies*, 56, pp.163-173.
- Double, R., 1991, 'Determinism and the Experience of Freedom', *Pacific Philosophical Quarterly*, 72, pp.1-8.
- Double, R., 1991, **The Non-Reality of Free Will** (Oxford: Oxford University Press.)
- Eccles, J.C., 1986, 'Do Mental Events Cause Neural Events Analogously to the Probability Fields of Quantum Mechanics?', *Proc. R. Soc. Lond., B* 227, pp.411-428.
- Eccles, J.C., 1989, **Evolution of the Self: Creation of the Self** (London: Routledge.)
- Edwards, P., 1961, 'Hard and Soft Determinism', in Hook.
- d'Espagnat, B., 1989, **Reality and the Physicist** (Cambridge: CUP.)
- Fischer, J.M., ed., 1986, **Moral Responsibility** (Ithaca, NY: Cornell University Press.)
- Flood, R. & Lockwood, M., eds., 1986, **The Nature of Time** (Oxford: Basil Blackwell.)
- Fouts, R.S. & Fouts, D.H., 1993, 'Chimpanzees' Use of Sign Language', in Cavalieri & Singer.
- Frankfurt, H.G., 1969, 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66, pp.829-839.
- Frankfurt, H.G., 1971, 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy*, 68, pp.5-20.
- Frankfurt, H.G., 1975, 'Three Concepts of Free Action', *Aristotelian Society Proceedings Supplementary*, Volume 49, pp.113-25.
- Gardner, R.A., Gardner, B.T., & Van Cantfort, T.E., 1989, **Teaching Sign Language to Chimpanzees** (Albany: State University of New York Press.)
- Gomberg, P., 1978, 'Free Will as Ultimate Responsibility', *American Philosophical Quarterly*, 15, pp.205-211.

- Goodall, J., 1986, **The Chimpanzees of Gombe - Patterns of Behaviour** (London: The Belknap Press of Harvard University Press.)
- Gregory, R.L., ed., 1987, **The Oxford Companion to the Mind** (Oxford: OUP.)
- Hawking, S.W., 1988, **A Brief History of Time** (London: Bantam Press.)
- Heinaman, R., 1986, 'Incompatibilism without the Principle of Alternate Possibilities', *Australasian Journal of Philosophy*, 64, pp.266-276.
- Hook, S., (ed), 1961, **Determinism and Freedom in the Age of Modern Science** (New York: Collier.)
- Honderich, ed., 1973, **Essays on Freedom of Action** (London: Routledge & Kegan Paul.)
- Honderich, T., 1988a, **Mind and Brain** (Oxford: Clarendon Press.)
- Honderich, T., 1988b, **The Consequences of Determinism** (Oxford: Clarendon Press.)
- Honderich, T., 1991, 'Better the Union Theory', *Analysis*, 51, pp.166-173.
- Honderich, T., 1993, **How Free Are You?** (Oxford: Oxford University Press.)
- Honderich, T. & Burnyeat, M., 1979, **Philosophy As It Is** (London: Penguin Books.)
- Kane, R., 1985, **Free Will and Values** (Albany: SUNY Press.)
- Kapitan, T, 1994, 'Double's "The Non-Reality of Free Will"', *Nous*, 28, pp.90-95.
- Kendall, R.E., 1987, 'Schizophrenia', in Gregory.
- Kenny, A., 1978, **Freewill and Responsibility** (London: Routledge & Kegan Paul.)
- Klein, M., 1990, **Determinism, Blameworthiness, and Deprivation** (Oxford: Clarendon Press.)
- Larmer, R.A., 1986, 'Free Will, Hegemony and Neurophysiological Indeterminism', *Philosophia*, 16, pp.177-189.
- Leakey, R & Lewin, R., 1992, **Origins Reconsidered** (London: Little, Brown and Company.)
- Lehrer, K., (ed), 1966, **Freedom and Determinism** (New York: Random House.)
- Levin, M., 1979, **Metaphysics and the Mind-Body Problem** (Oxford: OUP.)

- Lord, C., Ross, L., and Lepper, M.R., 1979, 'Biased Assimilation and Attitude Polarization', *Journal of Personality and Social Psychology*, 37, pp.2098-2109.
- Lucas, J.R., 1986, 'The Open Future', in Flood & Lockwood.
- Lucas, J.R., 1993, **Responsibility** (Oxford: Clarendon Press.)
- Lycan, W., 1987, **Consciousness** (Cambridge, MA: The MIT Press.)
- MacKay, D.M., 1960, 'On the Logical Indeterminacy of a Free Choice', *Mind*, 69, pp.31-40.
- Maxwell, N., 1985, 'Are Probabilism and Special Relativity Incompatible?', *Philosophy of Science*, 52, pp.23-43.
- Maxwell, N., 1988, 'Quantum Propensiton Theory: A Testable Resolution of the Wave/Particle Dilemma', *British Journal for the Philosophy of Science*, 39, pp.1-50.
- Moore, G.E., 1965 (1912), **Ethics** (OUP, New York.)
- Nisbett, R. & Ross, L., 1980, **Human Interference: Strategies and Shortcomings of Social Judgement** (Englewood Cliffs: Prentice Hall.)
- Popper, K.R., 1950, 'Indeterminism in Quantum Physics and Classical Physics', *British Journal for the Philosophy of Science*, 1, Part I pp.117-133, Part II pp.173-195.
- Popper, K.R., 1982, **The Open Universe: An Argument for Indeterminism**, volume II of **The Postscript** (London: Hutchinson.)
- Popper, K.R. & Eccles, J.C., 1983, **The Self and Its Brain** (London: Routledge & Kegan Paul.)
- Schlossberger, E., 1986, 'Why We Are Responsible for Our Emotions', *Mind*, 95, pp.37-56.
- Schoeman, F., (ed), 1987, **Responsibility, Character and the Emotions** (Cambridge, CUP.)
- Sher, G., 1987, **Desert** (Princeton, NJ: Princeton University Press.)
- Smith, O.R. & Jones, P., 1986, **The Philosophy of Mind** (Cambridge: CUP.)
- Stannard, R., 1993, **Doing Away With God?** (London: Marshall Pickering.)
- Strasser, M., 1988, 'Frankfurt, Aristotle and PAP', *Southern Journal of Philosophy*, 26, pp.235-246
- Strawson, G., 1986, **Freedom and Belief** (Oxford: Clarendon Press.)

- Strawson, G., 1989, 'Consciousness, Free Will, and the Unimportance of Determinism', *Inquiry*, 32, pp.3-27.
- Strawson, P.F., 1980, 'Reply to Ayer and Bennett', in Van Straaten.
- Strawson, P.F., 1982 (1962), 'Freedom and Resentment', in Watson, 1982.
- Swinburne, R., 1986, *The Evolution of the Soul* (Oxford: Clarendon Press.)
- Taylor, R., 1966, *Action and Purpose* (Englewood Cliffs: Prentice-Hall.)
- Thornton, M., 1989, *Do We Have Free Will?* (Bristol: Bristol Classical Press.)
- Thorp, J., 1980, *Free Will: A defense Against Neurophysiological Determinism* (London: Routledge & Kegan Paul.)
- Trusted, J., 1984, *Free Will and Responsibility* (Oxford: OUP.)
- van Inwagen, P., 1983, *An Essay on Free Will* (Oxford: OUP.)
- Van Straaten, Z., 1980, ed., *Philosophical Subjects* (Oxford: Clarendon Press.)
- de Waal, F., 1982, *Chimpanzee Politics - Power and Sex Among Apes* (London: Unwin Paperbacks.)
- Waller, B.N., 1989a, 'Denying Moral Responsibility: the Difference it Makes', *Analysis*, 49, pp.44-47.
- Waller, B.N., 1989b, 'Uneven Starts and Just Deserts', *Analysis*, 49, pp.209-213.
- Waller, B.N., 1993, 'Responsibility and the Self-made Self', *Analysis*, 53, pp.45-51.
- Watson, G., 1975 'Free Agency', *Journal of Philosophy*, 72, pp.205-220.
- Watson, G., 1982, ed., *Free Will* (Oxford: Oxford University Press.)
- Watson, G., 1987a, 'Free Action and Free Will', *Mind*, 96, pp.145-172.
- Watson, G., 1987b, 'Responsibility and the Limits of Evil', in Schoeman.
- Weatherford, R., 1991, *The Implications of Determinism* (London: Routledge.)
- Wiggins, D., 1973, 'Towards a Reasonable Libertarianism' in Honderich, 1973.

Wolf, S., 1990, **Freedom Within Reason** (Oxford: Oxford University Press.)

Wolpert, L., 1992, **The Unnatural Nature of Science** (London: Faber and Faber.)