Free will and luck

Harrison, Gerald Kingsley

Free Will and Luck

Name: **Gerald Kingsley Harrison**

Degree for which thesis is being submitted: **Ph.D**

University: **University of Durham**

Department: **Philosophy Department**

Year of submission: **2005**

With thanks to the Arts and Humanities Research Board for the support I have received.

Free Will and Luck

Gerald Kingsley Harrison

Abstract

The problem of free will is a problem about control and luck. If causal determinism is true, then everything we do is ultimately a matter of luck, as it is if causal determinism is false. Either way we seem to lack free will of the kind needed for moral responsibility. In this thesis a case is built for a certain type of modest incompatibilist view on free will. It is argued that it makes no difference in terms of control whether determinism or indeterminism obtains. What matters is that we have a certain kind of ownership over what we do. Causal determinism rules this out, but indeterminism does not. This has the upshot that not only does free will turn out to be compatible with luck, exposure to a certain kind of luck is actually required, for unless we are exposed to this kind of luck our actions will not be truly ours. By providing luck with a positive role this thesis invites a re-evaluation of the reasons causal determinism destroys free will, and a re-evaluation of our attitudes towards luck. In short this thesis challenges the anti-luckism that lies behind the problem of free will.

Contents

Chapter 1

Introduction

Everything we do is ultimately a matter of luck. This does not preclude our possessing free will and being morally responsible for at least some of what we do, provided we are subject to luck in the right way. In fact, exposure to some luck is a requirement of free will. That, anyway, is the view that will be defended in what follows.

1.1 Moral Luck

My interest in the problem of free will came about after reading Nagel's famous article "Moral Luck" (1979).[1] In it, Nagel pointed out that everything we do seems ultimately to be attributable to factors outside our control. For instance, the kind of character that we have is a product of our heredity and upbringing, neither of which were our doing. The actions we perform, the decisions we make: these are just the inevitable consequence of our having a particular character or constitution in a particular environment. Just as our upbringing and heredity were not matters over which we exercised any control, nor was our environment. Furthermore, the actual impact our actions have upon the world is again a matter of luck. Whether my actions are successes or failures depends on a multitude of factors, most of which are outside my control. Nagel usefully outlined four ways in which we are subject to luck:

> One is the phenomenon of constitutive luck – the kind of person you are, where this is not just a question of what you deliberately do, but of your inclinations, capacities, and temperament. Another category is luck in one's circumstances – the kind of problems and situations one faces. The other two have to do with the causes and effects of action: luck in how one is determined by antecedent circumstances, and luck in the way one's actions and projects turn out. (1979, p. 28)

[1] Nagel's article was one of a pair on the subject, the other being by Bernard Williams. Nagel's paper was a response to William's, and both originally appeared in the *Proceedings of the Aristotelian Society,* Vol.

We can term these four kinds of luck: constitutive – luck in one's capacities and temperament; circumstantial – luck in one's circumstances; causal – luck in prior causes of one's action; and resultant – luck in the consequences of one's actions.

The problem is that if everything we do is ultimately a matter of the four kinds of luck Nagel outlined, this seems to threaten the idea that we can be morally responsible for what we do. How can it be appropriate to consider someone blameworthy or praiseworthy for something they have done if it turns out that their doing it was just a matter of luck? Moral responsibility (deserving praise or blame), presupposes that we have control over what we do. But if everything is ultimately a matter of luck, then this calls into question whether we really do have control, at least of the kind that could make us truly blameworthy or praiseworthy for what we do.

To see how natural and easy it is to end up with some startling and radical conclusions take an example of resultant luck. An assassin takes aim at the prime minister and shoots, intending to kill him. However, a strong gust of wind deflects the bullet, missing the prime minister. Should this fluke occurrence affect the assassin's blameworthiness? Would we say that the assassin would deserve more blame, and a more severe punishment, if his bullet had hit the prime minister? Why should a gust of wind affect the assassin's blameworthiness? After all, the assassin had absolutely nothing to do with whether or not the wind gusted. Put more dramatically, the gust of wind surely cannot have a bearing on which circle of hell he is flung to (Thomson 1993, p. 207). On reflection, there seems to be a strong case from luck for saying that whether the assassin is actually successful or not is irrelevant. It marks "nothing morally interesting about them" (Thomson 1993, p. 204). What matters is what the assassin intended to do, irrespective of whether he was successful or not in doing it. So, considerations of luck lead us to restrict the domain of moral responsibility to the internal world of intentions, decisions and so forth – the domain of the 'will'. Here at least, it is assumed, we have some kind of absolute or total control.

But luck can infect that internal world too. First, superficially, and then more fundamentally, as we shall see. Imagine the assassin again. He is deliberating about

---

50 (1976) and were subsequently reproduced with revisions in Williams (1981) and Nagel (1979) respectively.

whether or not to shoot the prime minister, but just as he is about to reach his decision, he is interrupted by a coughing fit, the occurrence of which was not under his control (Feinberg 1970). Now we seem committed to saying he is responsible for what he *would* have decided to do, had it not been for the occurrence of the coughing fit.[2] How can the coughing fit be allowed to affect which circle of hell he is flung to?

Note that by the same logic we would be committed to saying the assassin is responsible for *all* the decisions he would have made, had circumstances not under his control been different. For instance, maybe if he had not had the deprived upbringing he had, he would now be succouring the poor in a third world country. It was only by luck that he did *not* have such an upbringing, and so he deserves praise for all those laudable actions he would have performed if his luck had been different.

This is the problem of responsibility inflation. When we try and adjust for luck – when we try and neutralise its influence – we end up massively increasing both what we are blameworthy for (which now includes all those culpable decisions we would have made had the opportunity presented itself) and what we are praiseworthy for (which now includes all those laudable decisions that we would have made had the opportunity presented itself). What we actually do provides only one part of a much larger picture. It is on the basis of the larger picture that our true moral responsibility is determined.

We are going to encounter the problem of responsibility inflation later in the thesis. For whilst discussed by those interested in the problem of moral luck, it is largely overlooked by those who discuss the problem of free will. But at the moment we can put it to one side because the inflation problem only has bite if we can make sense of our being morally responsible at all.

The inflation problem presupposes that the assassin is responsible for their decision (or would have been had they made it). But what if we are subject to luck all the way through? What if the assassin's decision itself was a matter of luck in a more fundamental way than that outlined above? It is at this point that we enter the territory

[2] There are problems here regarding the truth of such counterfactuals – especially if we assume that an agent has genuine alternative possibilities at 'responsible' decision making moments. However I discuss this later – in Chapter 8 – and will ignore it for the present as it does not alter my main points here.

covered by the problem of free will. For the problem of free will suggests that our decisions themselves are as much a matter of our luck as gusts of wind and the like.

Here is how the problem of free will unfolds. Let us provisionally understand by the will our conscious decision making process. Now, as I noted above we normally take ourselves to have a special species of control over our decisions and choices. Our decisions are not things that *happen* to us, but things we do – we *make* decisions, as opposed to decisions occurring in us. And when we make decisions we typically take ourselves to have been able to decide otherwise. In other words, we take ourselves to have alternative possibilities.

The natural picture we have of ourselves as free agents is often expressed in terms of a metaphor of a garden of forking paths.[3] We take our future to be like such a garden – consisting not of *one* future path, but many possible pathways branching out ahead of us, and it is down to us, through our exercise of free will, which future we actualise. Indeed, so fundamental and natural does this connection between having alternative possibilities – an 'open' future – and free will seem to be that until very recently it was taken as virtually axiomatic, at least for the kind of free will needed for moral responsibility. Free will is *essentially* about having genuine alternative possibilities.

This intuitively attractive picture of free will has been widely endorsed by philosophers as disparate as Aristotle, Hume and Kant, and it continues to have many adherents today.[4] It is an idea that is conveniently encapsulated in the Principle of Alternative Possibilities:

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

---

[3] I believe the 'Garden of Forking Paths' metaphor is taken from Jorge Luis Borges short story by the same name, and has become the common metaphor in the contemporary literature on free will thanks to John Martin Fischer (Borges 1974, pp. 81-92; Fischer 1994).

[4] Aristotle, for instance, says that "where it is in our power to act, it is also in our power not to act" (1976 p. 122 [1113b1-14]) Kant says that for free will "the act as well as its opposite must be within the power of the subject at the moment of its taking place" (1960, p. 45). It would take too long to list all contemporary advocates of the garden of forking paths view, but some of the most prominent include Ginet (1990); Kane (1996); van Inwagen (1983); Ekstrom (2000).

The problem is that a certain thesis about the structural underpinnings of the universe – known as causal determinism – seems to threaten this intuitive picture of free will. Matters get worse. For on reflection, it appears that matters are no better if causal determinism is false. Either way a strong case can be made for saying that everything we do, is ultimately out of our control and as such a matter of luck in a way that undermines the propriety of many of our attitudes towards ourselves and others. This is how the problem unfolds.

It is possible that the thesis known as causal determinism - the claim that all the facts about the past, in conjunction with all the laws of nature, entail one unique future - is true.[5] If this turns out to be the case then it would seem that we do not have genuine alternative possibilities. Whatever actions we perform, whatever decisions we make, there was never any possibility of our having decided otherwise. The garden of forking paths, turns out to be just one path with no junctions at any point. Our choices are just intermediate events in a long sequence that stretches back to the big bang. All we do is just the inevitable result of our programming – programming that we had no hand in writing. In short, if causal determinism is true, then we lack ultimate control, and so everything we do turns out to be a matter of our luck.

1.2 The Problem of Free Will: Indeterminism

If causal determinism is false – so if indeterminism is true - then we might have genuine alternative possibilities and so an open future. For if indeterminism is true then it would be possible that, given exactly the same past and the laws of nature, we could have decided otherwise on at least some occasions. But it is hard to see exactly how indeterminism improves anything where control is concerned. For example, imagine that Larry is torn between doing what he thinks he morally ought to do, and what he really wants to do. On the one hand he judges that he ought to phone his ill friend David, but what he really wants to do is take a long hot soak in a bath. In the actual sequence Larry decides to have a bath, but because his decision making process was

[5] The most widely accepted interpretations of quantum mechanics assume the world to be indeterministic, but this indeterminacy seems only to be present at the micro level and cancels out at the macro level. This may or may not be the case where our neural events are concerned – so it does, at the moment, remain an open question whether our decisions and thought processes are indeterministic. But there certainly is a very good chance that determinism – or something sufficiently close to determinism – obtains for our neural events.

indeterministic he could have decided to phone David right up to the moment of the decision, given exactly the same laws of nature and past circumstances. But then it looks as if it was just a matter of luck that Larry decided to have a bath rather than phone David. For there is nothing about Larry's powers, capacities, states of mind, character – anything at all – that explains why Larry decided one way rather than the other. If, for instance, God were to take Larry to one side and rewind the clock to the moment just prior to his decision and re-run the sequence a number of times, sometimes Larry would decide one way, and sometimes the other. And there would be no explanation of this, beyond "dumb luck" (Mele 1999b and 1998, pp. 582-583). So whilst the denial of determinism – indeterminism – can provide us with alternative possibilities and so make it the case that it is no longer inevitable that we make the decisions that we do, it seems to do nothing to prevent all that we do being a matter of luck. In fact it may even make matters worse. If our decisions are just random, then we seem to have *less* control over them than we would if determinism were true.

1.3 The Two Camps: Compatibilism

There have been many reactions to the problem of free will and it is "perhaps the most voluminously debated of all philosophical problems" (Matson 1987, p. 158). This continues to be the case with a truly huge amount having been written just in the past forty years with so many new developments and variations on old positions that it has become increasingly difficult to keep track (Kane 2002, p. 3). Nevertheless, we can distinguish two broad camps. There are those who claim that causal determinism does not rule out free will and moral responsibility. Those in this camp are called compatibilists. There are those who hold that determinism does rule out free will and responsibility. These are the incompatibilists. There are divisions within these camps, and there are small bands who do not fall into either camp, but we will mention these others later.

The compatibilist has the burden of explaining just how we can have adequate control over what we do if determinism is true. The incompatibilist has the burden of explaining how indeterminism can provide us with anything in terms of control. Both sides have employed various strategies to cope with their respective problems.

Historically, the compatibilists have tended to favour a particular reading of PAP – a conditional reading which renders it consistent with the truth of causal determinism. This kind of compatibilist tries to show how, viewed a certain way, we can reconcile the garden of forking paths picture with determinism. As such determinism should not be seen as the threat it appears.

But more recently there has emerged a new breed of compatibilist who rejects PAP altogether, and so rejects the garden of forking paths picture of free will, or at least argues that this is not the kind of free will needed for moral responsibility. In effect this breed of compatibilist accepts that ultimately all that we do is a matter of luck, but points out that this is consistent with our exercising a robust type of control over what we do, sufficient for free will of a kind relevant to our being responsible.

1.4 The Two Camps: Incompatibilism

The incompatibilist will typically affirm the garden of forking paths picture. To deal with the problem of luck they will often bring in some kind of 'extra factor' – a noumenal self, a soul, or more typically of modern incompatibilists, a distinct kind of occurrent causation - to deliver path-picking control. Possession of this type of control prevents all that we do being a matter of luck, and as such we have the kind of control sufficient for responsibility grounding free will.

However, it is mysterious exactly how these 'extra factors' deliver the control in question, so increasingly we find incompatibilists trying to do without them. For part of what the garden of forking paths picture encapsulates is the thought that free will involves being the initiating or ultimate source of one's actions, rather than being a transitional link in an extended deterministic chain. Satisfying this demand does not seem to require additional control of the kind that 'extra factors' deliver. If it is indeterministic which way we will consciously choose to decide then we are free from domination by the past and so what we do can be said to be in some sense truly up to us in a way that would not be the case if casual determinism were true. So there is room to argue that it is enough that we are the ultimate sources or initiators of our actions for us

to have the kind of free will needed for moral responsibility.[6] I call this kind of incompatibilist a 'source' incompatibilist, in contrast with the previous 'control' incompatibilist to denote the difference in focus.

Because the source incompatibilist does not bring in any extra factors to deliver path-picking control, ultimate luck remains. This kind of incompatibilist is therefore relevantly analogous to certain compatibilist positions insofar as both have to tacitly acknowledge that ultimately, all that we do is a matter of luck. So both source incompatibilists and compatibilists share the burden of reconciling free will and responsibility with ultimate luck.

1.5 Primary Goals and Chapter Synopsis

In this thesis I will be defending a source incompatibilist view of the kind that Robert Kane has done much to develop. But my first goal is to bring to light the real weakness of compatibilist accounts of free will. Whilst the tendency has been to focus on the supposed inadequacy of compatibilist control, and the way in which it exposes us to ultimate luck, I argue that the most telling problem facing the compatibilist is showing how we can have ownership over what we do. This problem is brought out by the difficulty compatibilists have in dealing with cases involving clandestine manipulation.

This first goal is played out in chapters 2, 3 and 4. In Chapter 2 I outline a number of different compatibilist approaches to free will, and I make some standard criticisms and go on to outline the reason-responsive view of those compatibilists who reject PAP altogether. In Chapter 3 I turn my attention to the basis upon which PAP is being rejected – the Frankfurt-style case. I argue that standard Frankfurt-style cases do not work, but I do allow that a variation developed by Fischer might *half* work. It provides some evidence that path-picking control of the sort commonly thought to require the introduction of extra factors is not needed for responsibility grounding free will. But the case provides no counterexample to PAP. As such it invites us to question the reason PAP is associated with free will and moral responsibility.

---

[6] This kind of 'modest' incompatibilist position has been suggested by Wiggins (2003); Nozick (1981); and Mele, and has been extensively developed by Kane (1996), and variations defended by Balaguar

Then in Chapter 4 I turn my attention to cases involving covert non-constraining control – or CNC control as it has become known. I follow others in arguing that there is no relevant difference between being subject to CNC control and acting under causal determinism, and because the former subverts responsibility so too does the latter. I contend that what CNC control cases highlight is our lack of adequate ownership over what we do under determinism.

My second goal is to be as clear and honest about the kind of source incompatibilist position I defend, in a way that others who defend this position have not been. Rather than being coy or evasive about exposure to ultimate luck, I encourage honest acceptance that in the absence of extra factors, source incompatibilist positions leave the agent exposed to ultimate luck. The attempts by some source incompatibilists to deny exposure to ultimate luck all fail, and display a background 'anti-luckism' which I hold actually prevents the real merit of the view from being seen.

This second goal is pursued in chapters 5 and 6. In Chapter 5 I argue that incompatibilist control is either incoherent or ineliminably mysterious, and that either way it is a good idea to see how far one can go without it. I outline two sorts of 'modest' incompatibilist position – modest because they do not bring in any exotic extra factors, but make do with indeterministic event causation. I argue that neither the Valerian variety of modest incompatibilism nor the action centred variety associated with Robert Kane provide any enhancement of control, and as such these views leave one as exposed to ultimate luck as one would be under determinism. I consider Kane's detailed responses to this kind of charge, and find them all wanting.

In Chapter 6 I consider whether there is a case for saying that indeterminism, as it is introduced under either Valerian or Kanean modest incompatibilist views, erodes control. I answer no. We are neither better off, nor worse off, in terms of control than we would be under determinism.

My third goal is to highlight how genuine alternative possibilities provide us with the kind of ownership over what we do that had previously been shown to be lacking under determinism. In this way I show how exposure to certain sorts of luck actually delivers

---

(2004) and Ekstrom (2000).

something of value, and as such we can start to see our exposure to luck in a more positive way. Source incompatibilism is thereby shown to be a more attractive position than its compatibilist rivals, both providing a depth of ownership unavailable under determinism and vindicating our intuitive picture of free will as involving an open future of genuine alternative possibilities.

At the same time I also show the deficiencies of certain forms of source incompatibilism. This goal is pursued in Chapter 7, where I reject Valerian incompatibilism in favour of Kane's modest incompatibilism. I also consider a number of objections to my favoured position, and make refinements as necessary. I end the chapter by highlighting that given what has been argued so far, the modest incompatibilist position is to be preferred to its compatibilist rivals.

My fourth goal is to show the most fruitful way source incompatibilism can be shown to its full advantage against rival control incompatibilist positions. Whilst the problem of free will can be understood as a problem to do with luck and control, there is another problem to do with luck which has been virtually entirely neglected by free will theorists. For the problem of free will is a problem to do with trying to get moral responsibility off the ground so to speak – a problem trying to show how we can have adequate control over what we actually do. But getting responsibility off the ground is solves only half the problem. There is then a problem trying to restrict responsibility to those actions we actually perform, and not all those we could have performed had factors not under our control been different. This second problem is particularly hazardous for those who hold fundamentally 'anti-luckist' positions of the kind motivating the need for path-picking control. The source incompatibilist faces no similar difficulties here.

A related goal involves drawing attention to an alternative story behind the supposed need for incompatibilist control in contrast to the usual 'anti luckist' rationale. The story I propose sees incompatibilist control in a restorative capacity, with the main focus on the ownership rather than control. Reflection on the implications of what consistent 'anti-luckism' would commit one to, combined with the recognition that incompatibilist-control would still leave us heavily exposed to luck, means that the anti-luckist rationale behind incompatibilist control should be dropped, and that the rationale

that I propose is preferable. But once this is accepted, then source incompatibilism appears a leaner, more attractive position than control-incompatibilism. These two goals are pursued in Chapter 8.

Finally, having shown how the modest incompatibilist position can outshine its rivals I will use Chapter 9 to draw together the main themes of the previous chapters, and highlight both the limits and the helpfulness of what has been argued.

Chapter 2

Compatibilist Strategies

In this chapter I will argue that causal determinism threatens the idea that we have genuine alternative possibilities. I will present a more detailed version of this argument before going on to consider various compatibilist responses. I raise familiar objections to some of the proposals. Finally I sketch the kind of compatibilism advocated by the most recent breed of compatibilists, namely those who reject the need for alternative possibilities altogether. Critical assessment of the argument upon which they base their rejection of PAP will wait until the next chapter.

2.1 The Consequence Argument

In the previous chapter I suggested that causal determinism, if true, would rule out alternative possibilities. I will now make that argument in more detail. It has become known as the 'The Consequence Argument' and is given its general form in the following often quoted passage:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (van Inwagen 1983, p. 16)

There are a number of different versions of the argument that can be extracted from the above passage, but they are all essentially the same insofar as if one fails, then they all do (van Inwagen 1983, p. 57; for objections see Fischer 1994, pp. 248-249 n. 12). The version that I will outline below is one given by Fischer in his 1994.[7] It depends upon three very plausible principles and the assumption of determinism. The first Fischer calls the Principle of the Transfer of Powerlessness:

[7] Others who have made versions of this argument include: van Inwagen 1983, pp. 55-105; Ginet 1966 pp. 87-104; Ginet 1983; Ginet 1990; Wiggins 2003; and in a theological context, Pike 1965.

> [I]f a person is powerless over one thing, and powerless over that thing's leading to another, then the person is powerless over the second thing.... Suppose, for example, that a meteorite hits my house today, and that I am powerless to prevent this. Imagine, further, that the meteorite is being enough to destroy my roof, so that it is true that if it hits my house, it destroys my roof (and, again, I am powerless to alter this truth). It seems that it follows that the meteorite destroys my roof, and that I am powerless to prevent the destruction of my roof. (Fischer 1994, pp. 8-9)

The second principle Fischer terms the Principle of the Fixity of the Past. This principle says that the past is now fixed and out of our control. I cannot, for instance, now change the fact that I had cornflakes for breakfast this morning. As such, I do not have the power to do something which would require the past to be different. As Fischer puts it "the principle says that if a person's performing a certain action would require some actual fact about the past *not* to have been a fact, then the person *cannot* perform the act" (1994, p. 9). The third principle says something equally common-sensical about the laws of nature, and can be termed the Principle of the Fixity of the Laws (1994, p. 9). For instance, it is a law of nature – or seems to be – that nothing can travel faster than the speed of light. If that is a law of nature, then there is nothing that I can do that is contrary to that law. The laws of nature, like the past, are out of our control.

The fourth and final ingredient is the thesis of causal determinism. So, at any given time, a complete statement of the facts about the past, and a complete statement of the laws of nature, entails all future truths. With these ingredients in place, Fischer's version of the Consequence Argument can be presented as follows. Agent S, performs some act X, which we can take to be a mental action such as making a decision at time t2:

> [I]f determinism is true, and s1 is the total state of the world at t1, one of the following conditionals must be true:

1. If S were to refrain from doing X at t2, it would not have been the total state of the world at t1.

2. If S were to refrain from doing X at t2, then some natural law which actually obtains would not obtain.
3. If X were to refrain from doing X at t2, then either is would not have been the total state of the world at it, or some natural law which actually obtains would not obtain.

> But if (1) is true, then... [assuming the truth of the Principle of the Fixity of the Past] S cannot refrain from doing X at t2. Similarly, if (2) is true, then... [assuming the truth of the Principle of the Fixity of the Laws] S cannot refrain from doing X at t2. Finally, if (1)'s truth implies that S cannot refrain from ding x at t2, and (2)'s truth implies that S cannot refrain from doing X at t2, then it follows that if (3) is true, S cannot refrain from doing X at t2. The conclusion of this argument is that if determinism is true, then S cannot do anything other than what he actually does at t2. Generalising this result,... if determinism is true, none of us is free to do other than what he actually does. (Fischer 1994, pp. 62-63)

All the ingredients of this argument are plausible, common-sense principles, and furthermore determinism is a thesis about the structural underpinnings of the universe which may very well turn out to be true. Given in addition the intuitive plausibility of PAP this argument appears devastating to free will and moral responsibility.

What, then, does the compatibilist say to this apparently devastating argument? Well, some acknowledge that we do need genuine alternative possibilities, but they deny that determinism actually rules them out. Others argue that we should understand the intuitive demand for an 'ability to do otherwise' conditionally. When we interpret the demand for an ability to do otherwise in this way, then it turns out to be compatible with determinism. In other words, these compatibilists accept that we do not have genuine alternative possibilities, but argue that we still have an ability to do otherwise in a sense relevant to free will and responsibility. Others try to diffuse the problem of free will by arguing that we have over-intellectualised it, and that in fact our practices of blaming and praising are insulated against discoveries about the structural underpinnings of the universe (and that there is something faintly absurd about thinking that such discoveries could really make a difference). But the most popular recent strategy by compatibilists

has been to reject that we need alternative possibilities and the associated ability to decide otherwise. Whilst we may intuitively feel that alternative possibilities are needed, these compatibilists argue that a certain kind of case – known as a Frankfurt-style case – shows this assumption to be mistaken, at least with respect to the kind of free will and control needed for responsibility. PAP, in other words, is false and the only kind of control presupposed by moral responsibility and free will is of a kind compatible with determinism.

It would be impossible to do full justice to all the various compatibilist arguments that have been offered with respect to the above selection of positions. Some of the above are therefore going to have to be given rather short thrift. Two of the three positions above will be given some time in court. No doubt not enough time to fully do them justice, but enough, I hope, to highlight their deficiencies. The first position that is going to be given the most cursory treatment.

2.2 Rejecting the Principles

There are compatibilists who believe that we *do* have genuine alternative possibilities, even if determinism is true. But in maintaining this, they have to challenge one of the above principles, either the fixity of the past, the fixity of the laws, or the transfer principle. Typically, those who challenge the fixity of the past or the fixity of the laws are actually objecting to the sense of 'could have done otherwise' that the consequence argument seeks to show determinism deprives us of.[8] In other words, these compatibilists are really of the second sort – they are advocating a conditional analysis of control.[9] But those who question the transfer principle leave open, if their arguments work, that we could have genuine alternative possibilities even if determinism is true. This is because if the transfer principle is false, then the Consequence Argument does not establish its conclusion. That does not mean that we actually *do* have genuine alternative possibilities if determinism is true, but it does mean that it cannot be

[8] Those who have made such arguments include: Gallois 1977; Horgan 1985; Lewis 2003; Narveson 1977; Saunders 1968; Vihvelin 1991. It is also possible for the compatibilist to question the nature of causation and to appeal to a non-necessitarian view of the laws of nature. Hume is famous for advocating a non-necessitarian view of the laws. According to this view the same pasts are consistent with different laws, and as such the past being what it is, does not entail that there is just one future. See for instance, Berofsky 1987 chs. 8 and 9.

[9] One does not *have* to do this – for one could just out and out reject either the principle of the fixity of the past or the fixity of the laws. But then that would be an extremely controversial move.

uncontroversially *demonstrated* to be the case.

Although the arguments over the validity of the transfer principle are by no means settled, I have nothing to add to the discussion and in my view the transfer principle is sufficiently common-sensical to make any position which depends upon rejecting it, a highly controversial position.[10] Until matters are settled here, I believe the burden of proof lies with those who would reject the Transfer Principle, and so will simply assume that the consequence argument works. The Consequence Argument makes it highly plausible that determinism is incompatible with genuine alternative possibilities and we have no non-controversial reason to think otherwise. That, I believe, is as much as one can expect in this area. From now on I will assume that determinism really does rule out genuine alternative possibilities. Even if I am wrong about this, it will be instructive to explore what the implications of lacking genuine alternative possibilities actually are.

2.3 Conditional Analysis of Ability

The next kind of compatibilist accepts that we do not have genuine alternative possibilities of the kind ruled out above. But they reject that this means we lack the ability to do otherwise. This is because these compatibilists give a conditional analysis of what it means to say that someone could have done otherwise, and the conditional analysis is consistent with determinism being true.

Consider first if determinism is true, then no one has he ability to do otherwise when the past and the laws are held fixed. However, what this kind of compatibilist claims is that we do not need *this* kind of ability in order to have free will and moral responsibility. Rather the sense of 'ability to do otherwise' relevant to free will and responsibility is the ability to do otherwise if one chooses to do otherwise (Hume 1955, p. 73; Ayer 1954; Hobart 1934; Moore 1993, pp. 299-313). Hobbes was amongst the first to suggest such an analysis when he said that "he is free to do a thing, that may do it if he have the will to do it, and may forbear, if he have the will to forbear" (Hobbes 1962, p. 240). Hume also gave a conditional analysis of freedom:

> [I]f we choose to remain at rest, we may; if we choose to move, we also may. Now this hypothetical liberty is universally allowed to belong to everyone who is not a prisoner and in chains. (1955, p. 104)

Mill propounded such a view in the nineteenth century and G. E. Moore took the conditional analysis into the twentieth century, where it was also taken up by Ayer (1954), Schlick (1966), Hobart (1934), and Davidson (1973) amongst others. It continues to be defended in one form or another by Bok (1998), Kapitan (2002), Vivhelin (1991), and Peacocke (1999).

To get clearer about the conditional analysis (and ignoring differences in detail between all of the above) consider the example of a television remote control. We might say of a remote control that it changed the channel to BBC1 because the BBC1 button on its keypad was pressed. But it *would* have changed channels to BBC2 *if* the BBC2 button had been pressed. The remote control is a control mechanism due to this fact about it. We would say of someone who had possession of this remote control, that though in the actual sequence they changed channels to BBC1, they had the ability to change channel to BBC2, because if they had pressed the BBC2 button the channel would have changed accordingly. The compatibilist is saying that our decisions are exhibit control in the same way: we decide one way in the actual sequence, but we *could* have decided otherwise, *if* we had wanted. Just as the truth of determinism would not in any way render the television remote control less of a control mechanism, similarly the truth of determinism would not threaten our control either. Just as the remote control is suitably responsive to its keypad, so we are suitably responsive to our wants or reasons.

I think we should admit that this is an analysis of a kind of control, and the control in question is compatible with determinism. After all, we clearly would still have some species of control over what we do even if determinism is true (if determinism were announced to be true tomorrow, would we feel out of control? Of course not). We should also concede that very often it is precisely this conditional sense of could have done otherwise that we invoke. I might say 'oh, well I could have done that if I'd

---

[10] Those who have challenged the Transfer Principle include: Widerker 1987; McKay and Johnson 1996; Slote 1982. And for replies to these challenges see: Fischer 1994, pp. 29-44; O'Connor 2000 pp. 3-15; Finch and Warfield 1998.

wanted' –here I am clearly invoking the conditional sense of 'could'. I am reporting that I have a certain ability.

But there are problems when it comes to making the case for saying that this is the kind of ability to do otherwise associated with free will and responsibility. Firstly, it seems to deliver the wrong verdict in a number of cases. For instance, if Jack cannot go outside due to his agoraphobia, it might nevertheless be true that Jack would go outside if he chose to. Granted, he is not going to choose t – because he has agoraphobia! But that does not alter the fact that he would if he chose to. In this case the conditional analysis delivers the verdict that Jack has the ability to go outside. Yet, Jack clearly was not *free* to go outside in any sense relevant to free will and responsibility (on this see Lehrer 1966, and for discussion Berofsky 2002).

More devastatingly, the conditional sense of ability just does not seem to be what we are talking about when we are talking about the kind of alternative possibilities associated with moral responsibility. When we say that we have the ability to decide otherwise, we sometimes mean that we have the ability to decide otherwise *given conditions exactly as they are* - what Austin referred to as the 'all in' sense of can (1961).[11] This is especially so when it comes to cases involving blame. When we blame someone for making some morally reprehensible decision we assume that they could have decided otherwise. But we assume here that they could have decided otherwise *in these exact circumstances,* rather than in different circumstances. It seems entirely irrelevant what she would do in different circumstances.

Turning again to the remote control, it can be said to have controlled the channel change, but it did not have control over *how* it controlled the channel change. For that was a matter determined by whichever of its buttons had been pressed. It did not have control over how it exercised control in other words. Likewise we might control our decisions in virtue of the fact that we could have decided otherwise if he had wanted to. But we lack control over our 'wants' in the same way that the remote control does not

[11] Austin gives a now famous example in a footnote to his paper *Ifs and Cans*:

> Consider the case where I miss a very short putt and kick myself because I could have holed it. It is not that I should have holed it if I had tried: I did try, and missed. It is not that I should have holed it if conditions had been different: that might of course be so, but I am talking

control how its buttons are pressed. It is this kind of reflection that leads back to the concern about luck. If we do not *ultimately* control how we decide in any actual sequence, then it looks as if it is just a matter of their luck that we found ourselves so constituted that on *this* particular occasion it was inevitable that we would make the decision we did.

2.4 Hilary Bok

Hilary Bok is a contemporary compatibilist who defends the conditional analysis, and it will be instructive to see what she says at this point. Bok holds that it is the conditional understanding of ability that we are committed to when we deliberate, and so the conditional sense of ability to do otherwise is the one relevant to free will and responsibility. To get clearer, let us first distinguish between theoretical reason and practical reason. When we employ theoretical reason we are trying to determine what is the case – that is the purpose of theoretical reasoning. But the purpose of practical reason is to find out what to do. (Practical reasoning can yield two kinds of judgement – a judgement about what it would be best, or what one ought to do – normative judgements – and actual choices or decisions (Bok 1998 ch. 2; Kane 1996, p. 21)). Depending upon which perspective we are adopting – the theoretical or the practical – we have different ends in mind. What Bok argues is that from the theoretical perspective the compatibilists conditional analysis of control does look empty and irrelevant. When we adopt the theoretical perspective our decisions are just the inevitable output of mechanical processes. But when we adopt the perspective of practical reason – so when we are trying to determine what to do – then the compatibilist analysis appears correct, or at least it is the compatibilist's notion of possibility that we entertain. When we deliberate we take our alternatives to be conditional alternatives. From the practical perspective, the type of freedom we have to assume we have is the compatibilist conditional freedom to decide otherwise. As Bok states:

> From the practical point of view... our use of [the compatibilist conditional] conception of possibility, as opposed to the narrower possibility *tout court* [the incompatibilist conception of possibility – something is possible only if it is

---

about conditions as they precisely were, and asserting that I could have holed it. There is the rub. (1961, p. 218)

> possible holding fixed the past and the laws of nature], is both unavoidable and rational. It is unavoidable because while we deliberate we cannot possible employ a conception of the alternatives that are available to us that is narrower than the set of actions that we would perform were we to choose to do so. It is rational because, for the purposes of deliberation, we must regard that question as open, we should not regard the various actions that we would perform if we chose as differing with respect to their possibility, since any of them would be possible *tout court* if we chose to perform it. Moreover, to determine whether or not a particular action is one which we would perform if we chose is to determine whether or not we can regard it as a possible object of choice: an action about which the question whether or not we have reason to perform it can legitimately be raised. (Bok 2003, p. 161)

When we are employing our practical reasoning, we are trying to determine which amongst a set of possible decisions we should make. Those possibilities that we will take as real for us, are those which, if we were to choose them, we could realise. In other words, I do not deliberate about whether or not to flutter around the room, because I take it that it would not be possible for me to flutter around the room *even if I chose to*. We seek through practical reasoning, to narrow our range of possibilities to those things that we could do if we chose to. Therefore, it seems that when we adopt the practical perspective, which is to say, when we are actually trying to determine what to do, we need only employ the compatibilist conditional conception of what it is possible that we could do. Bok then argues that this is the conception of possibility relevant to free will and responsibility.

However, even if Bok is right with respect to the conception of possibility that we employ when we are trying to determine what to do, that does not mean that the compatibilist conditional analysis of the ability to do otherwise is the correct one when it comes to the question of whether we act freely and are morally responsible. Bok's point is relevantly identical to a point that Hobbes made in the seventeenth century about the prospect of God having pre-determined all that we do. Hobbes said that this may be the case, but given that we do not know what God has pre-determined us to do, we still have to deliberate about what to do in the normal way (Hobbes 1962, p. 174). In practical terms, the news that God is controlling us is irrelevant. Likewise, the news that

determinism is true is, from the practical perspective, irrelevant, for it does not in itself help me figure out what I should do. Yet the news that God has pre-determined all of our actions *is,* very plausibly, relevant to the question of responsibility and similarly so could be the truth of determinism.[12]

I am going to elaborate further upon the problems that compatibilists have concerning cases involving clandestine manipulation (of the sort that God might be guilty of) in Chapter 4. At the moment we can note that when it comes to assessing whether we have the kind of free will required for responsibility it is not obvious that the practical standpoint is the correct one to take. For we do not just want to know whether to blame someone, (or to hold ourselves blameworthy) we want to know whether they actually *are* blameworthy. As such Bok's focus on the practical perspective does not settle the matter of which analysis of possibility – the conditional or the unconditional – is the correct one with respect to responsibility.

2.5 Fake Compatibilists

For reasons already given, the conditional interpretation seems artificial when offered as an analysis of the kind of ability to do otherwise we associate with responsibility. It does articulate a kind of control, but it is a kind of control that leaves us exposed to ultimate luck. The kind of control conditional abilities confer on us is of a sort relevantly identical to that exhibited by television remote controls and the like.

However, what needs to be borne in mind is that many (though *not* Bok above) of the compatibilists who have offered the conditional analysis have a particular view about moral responsibility. It is this view about moral responsibility which explains why this kind of compatibilist control seems to be the only kind needed. The view about moral responsibility in question is consequentialist.

Consequentialist conceptions of responsibility are defended by 'compatibilists' such as Dennett, Mill, Sidgwick, Schlick, Stevenson, Smart, and Ayer. However, the consequentialist conception of moral responsibility is just *not* what we are talking about

[12] In the famous exchange between Bishop Bramhall and Hobbes this was precisely the point Bramhall made, arguing that if God created our characters, and if our characters caused us to behave as we do, then

when we wonder whether determinism threatens our free will and responsibility. For the consequentialist conception of moral responsibility does not presuppose free will *at all*. I charge then, that consequentialist 'compatibilists' are not actually compatibilists at all.

To get clearer about the consequentialist conception of responsibility, let us first take the Utilitarian as our standard bearer for consequentialism, and turn to Sidgwick for elucidation.

> [T]he determinist can give to the terms 'ill-desert' and 'responsibility' a signification which is not only clear and definite, but, from a utilitarian point of view, the only suitable meaning. In this view, if I affirm that A is responsible for a harmful act, I mean that it is right to punish him for it; primarily, in order that the fear of punishment may prevent him and others from committing similar acts in the future. (1963, pp. 71-73)

As Watson puts it, the idea is that praise and blame are just "ways of *grading people with a view to influencing their attitudes and conduct,* rather than *judging* them" (2003, p. 15). If this *is* the full story behind our practices of blaming and praising, punishing and rewarding, then the only kind of control that such practices seem to presuppose is the conditional compatibilist variety. The rationale goes, that if blame and praise are just tools – ways of getting people to change their behaviour – then it only makes sense to blame and praise those who exercise control over what they are doing, for only they will be responsive to such practices.[13] If determinism is true, then it would still make sense to try and alter people's behaviour by these methods, for determinism does not rule out our having conditional control. So, we get a connection between the appropriateness of blame and praise, and possession of control.[14] However, there are some well-known objections to this conception of responsibility, objections which highlight, amongst other things, how on this view the connection between having control and being responsible is not actually very strong.

---

responsibility for what we do traces to God and not us (1844, p. 30 and see Kane 1996, p. 35).

[13] When we come on to look at reason responsive compatibilist conceptions of control it will be apparent that the same rationale could be given. There really is no difference between the sort of control exercised, it is rather that a source compatibilists does not try to pretend that their conception of control is one that answers to what we normally take to be an ability to decide otherwise.

[14] Having a 'genuine' ability to do otherwise might actually be an obstacle to consequentialist goals. It would get in the way of *ensuring* that people only behave in certain ways, for a genuine ability to do otherwise introduces some uncertainty into the picture.

The most common objection to this consequentialist conception of responsibility is that it makes what one actually did of incidental importance demonstrated by the fact that a consequentialist rationale can be given for blaming someone who has done nothing wrong. This could come about for two main reasons – blaming you might have beneficial consequences on others (it may alter the behaviour of others, or, in the notorious 'scapegoating cases' it might allay other's fears or in some other way contribute towards social cohesion).[15] Blaming you might beneficially alter your future behaviour. In the latter case, imagine that you believe yourself to have done something wrong (in fact you just dreamt it). So long as you sincerely believe this, then blaming and punishing you as if you'd actually done wrong might be justified, on the grounds that if you go unpunished this might lead you to think that you can get away with wrongdoing. As long as you sincerely believe you did something wrong, it does not actually matter whether you actually did: the implications for your future behaviour are the same.

The problem is that if you accept this kind of rationale for our practices of blaming and praising, punishing and rewarding, you cannot at the same time be too concerned with what people actually do. In other words, you cannot be too concerned with whether those being blamed and praised are actually guilty of anything. The excuse 'but I didn't do it!' cannot settle the issue of an agent's blameworthiness (yet it is hard to think of a better excuse!). For the agent's blameworthiness is determined not by whether or not they actually did anything wrong, much less whether they did anything wrong freely, rather it is determined by whether it would be most utile to blame this person in the circumstances. That is why such a view about moral responsibility is radically at odds with our common sense conception, and why this kind of view just *isn't* compatibilist.[16] According to our common sense conception of moral responsibility, whether someone is truly blameworthy or praiseworthy is a factual matter in part determined by what the agent actually did, and whether or not the agent did it of their own free will where this includes having exercised appropriate control.[17] The problem with effect compatibilism,

---

[15] On the scapegoating problem see: McCloskey 1965, 1967 and 1973; Sprigge 1965; Smart 1973.

[16] There are other reasons to consider this conception of responsibility at odds with our 'common sense' conception, not least the fact that when we blame and praise others we do not normally take ourselves to be engaging in an exercise in behaviour modification (Squires 1969, p. 211).

[17] Smilansky makes this case too, and makes an interesting version of the 'punishing of the innocent' objection (2000, pp. 27-33; 1990b).

and why it is not really a form of compatibilism at all, is that it severs the intrinsic link between responsibility and free will and makes matters external to the agent determine their blameworthiness, namely whether or not it would be most utile to blame them. [18][19]

2.6 Strawsonian 'Reactive Naturalism'

A different compatibilist approach, but one which I hold shares some of the same faults as that above, is Peter Strawson's 'reactive naturalism' (I borrow the term from Smilansky 2000, p. 220). Strawson's view is expressed in his now famous, and much discussed "Freedom and Resentment" (1962). In that work Strawson introduces the useful term – reactive attitudes – to refer to attitudes such as gratitude, resentment, forgiveness and indignation. The reactive attitudes are "natural human reactions to the good or ill will or indifference of others towards us as displayed in *their* attitudes and actions" (Strawson 2003, p. 80) In the contemporary debate it is being an 'apt target for the reactive attitudes' – so being someone who is resent-worthy, forgivable, and so forth – is taken to be what being morally responsible *is*.

So far that is fine and settles nothing regarding the question of whether moral responsibility and free will require genuine alternative possibilities. However, what is novel about Strawson's approach is that he claims that it is a mistake to look for some kind of *deep* ground or justification for these attitudes and related practices. As Watson puts it "the... emphasis on abstract questions about the character of the laws of nature or of the causal structure of the world is... out of focus" (2003, pp. 15-16). For these

---

[18] One way to bring out the fact that effect compatibilism just *isn't* a form of compatibilism is the fact that it is perfectly possible for a hard incompatibilist to agree that consequential considerations now constitute the only basis upon which to praise and blame. Yet the hard incompatibilist, believes this precisely because they believe that we *do not* have free will. In other words, consequential justifications take over (and no doubt *alter* what we are and are not blame/punished for, or in light of) in the absence of a free will based justification.

I am *not* suggesting that the only basis upon which blame and punishment can be justified in the absence of free will are consequential grounds. There's no obvious reason why an absence of free will should mean that we lack basic rights, and these could put some limits on the extent to which consequential considerations determine how severely we are punished etc. See for instance: Pereboom 1995, 2001 and 2002 esp. pp. 479-480; Smilansky 2000, pp. 31-32).

[19] Above I have dismissed 'consequentialist' compatibilists as not really compatibilists at all. But in order to avoid misunderstanding let me say that what I have said above applies only to those who take a *purely* consequentialist approach to moral responsibility. It would not apply to those who believe, say, that it *is* on the basis of consequential considerations that we actively blame and praise etc, but that such practices need to be restricted according to a set of conditions – conditions to do with control and free will (so, it would not be appropriate to blame someone for something that they have not done, even if doing so would be utile).

attitudes, according to Strawson, do not stand in need of justification at all - they *constitute* moral responsibility. So, when we express our blame we are not voicing the recognition of some characteristic in the agent which itself justifies our expression of blame in any deep way, rather our expression of blame *just is* what it is to hold someone blameworthy. Again, our practice of actively punishing some and rewarding others, simply expresses our reactive attitudes, which do not themselves need justifying. And "[O]ur proneness to the reactive attitudes is a natural fact... neither calling for nor permitting a general rational justification" (P. Strawson 1980, p. 265). As Watson says:

> In Strawson's view, there is no such independent notion of responsibility that explains the propriety of the reactive attitudes. The explanatory priority is the other way around. It is not that we hold people responsible because they *are* responsible; rather, the idea (*our* idea) that we are responsible is to be understood by the practice, which itself is not a matter of holding some propositions to be true, but of expressing our concerns and demands about our treatment of one another. (1987, p. 258; see also Bennett 1980)

Consider that, as many compatibilists like to point out, most of the distinctions that we employ in everyday life to determine whether or not someone is blameworthy, are distinctions that would survive the revelation that determinism is true. For instance, when trying to determine blameworthiness we ask whether any normal excusing conditions apply – did the agent do what they did accidentally? By mistake? Inadvertently? Were they mad? If determinism is true this does not threaten the validity of any of *these* distinctions. If determinism is true we can still distinguish between the mad and the sane, we can still distinguish between the accidental and the intentional. It simply does not matter to the viability of these distinctions whether or not the agent had genuine alternative possibilities. It is he viability of these distinctions that is alone needed for our conception of responsibility to survive in a deterministic world. If determinism is true, then we still sometimes behave badly towards one another, we still sometimes intentionally step on one another's toes, and in these cases the reactive attitudes (blame, resentment and so forth) would still be 'apt'. In this way, moral responsibility remains viable even if determinism is true. Strawson goes on to say that,

if determinism were revealed to be true, we *would* not give up the reactive attitudes, nor should we, and nor could we if we tried.[20]

This view yields compatibilism insofar as, if correct, it really does not matter whether determinism is true or not. That we feel the reactive attitudes is a deep psychological fact about us, and is deeply embedded in our practices. Moral responsibility is, if one likes, completely insulated against discoveries such as determinism. As Strawson himself puts it:

> 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 inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question. (2003, p. 81)

But despite having been hugely influential, there are some serious problems with this kind of position. In the first place, one cannot simply cite the deep rootedness of our practices as in some way demonstrating that they do not presuppose that our world is a certain way. In everyday life we have to assume what philosophers can put in doubt – we just assume that, in the main, people have genuine alternative possibilities and adequate control over what they do. Granted, when we blame and praise people we do not normally enquire into whether they had *genuine* alternative possibilities – satisfying normal distinctions of the type that are not threatened by determinism is all that we do. But that does not mean that our practices and beliefs to not contain incompatibilist assumptions; it does not mean that *if* we found out that the agent in question lacked genuine alternative possibilities that we would not revise our judgement, that we would

[20] It is far from obvious that our reactive attitudes would remain unchanged in light of realising the absence of a desert base. If, on reflection, we come to see the determinism and the absence of genuine alternative possibilities as undermining our judgements of moral responsibility, then abandoning our attitudes, or at least beginning a process of revision, would seem possible or at least we could initiate a process of revision. After all, when we come to realise that a normal excusing condition obtains (that the person did not do what they did intentionally for instance) then we revise our attitude. So why be so sure that we cannot revise our attitudes when we come to see that excusing conditions obtain in every case (as an incompatibilist thinks would be the case if determinism turns out to be true)?

not, in other words, agree that blame is not appropriate. Why else do we even *have* a problem of free will if not because we do think that there is something threatening about determinism? It is not as if determinism is obviously benign. As Nagel puts it:

> When we first consider the possibility that all human actions are determined by heredity and environment, it threatens to defuse our reactive attitudes as effectively as does the information that a particular action was caused by the effects of a drug – despite all the differences between the two suppositions... Some of the externally imposed limitations... on our actions are evident to us. When we discover others, internal and less evident, our reactive attitudes toward the affected action tend to be defused... [it] seems no longer attributable in the required way to the person who must be the target of those attitudes. (2003, pp. 243-244)

P. Strawson's son is, interestingly, a critic of his father's position, and he too points out that our natural attitudes contain incompatibilist assumptions at their core.

> The fact that the incompatibilist intuition has such power for us is as much a natural fact about cogative beings like ourselves as is... our... commitment to the reactive attitudes... What is more, the roots of the incompatibilist intuition lie deep in the... reactive attitudes. (G. Strawson 1986, p. 88)

In other words, Strawson cannot just assume that our practices, and our beliefs about the appropriateness or aptness of the reactive attitudes do not contain, at their heart, incompatibilist assumptions about genuine alternative possibilities.

2.7 Strawson's Rationality Argument

Strawson claims that even if we do think that determinism subverts responsibility, and even if we did have a choice about whether or not to give up the reactive attitudes it would not be rational to do so. For the cost of giving up these attitudes, and the impoverished form of life that one would find oneself committed to, means that it simply would not make sense to give such attitudes up. Let us suppose that this is

correct.[21] One might say something similar with respect to someone who has structured their whole life around a particular religious view, and who happens upon a piece of evidence that shows their chosen religion to be a complete sham (perhaps they discover that their religious text is the work of a fraudster). Given just how centrally important this person's beliefs are, and given the extent to which their emotional and spiritual well-being is now intimately tied up with their particular religious view, we might judge that in this case the person has good reason to simply *ignore* the evidence in question, and to continue *as if* their religious text was actually the communication of a god.

Just as Kant said that we should act under a regulative principle to proceed *as if* God exists, so likewise all Strawson has shown the committed incompatibilist is that we have reason to proceed *as if* determinism were false, even if it turns out to be true. Deluding ourselves and entertaining wild and extravagant hopes can be rational if so much depends upon them.[22] But that does not tell us anything about what is in fact the case, and it does not tell us that our delusions do not, in fact, require that the world be a certain way – that there is not something that would vindicate our beliefs, and something else that would count against them.

Furthermore, whilst it might be rational to continue as if determinism were false, it might be equally rational to accept the truth and adjust – or try to adjust – one's attitudes accordingly. Again, Peter Strawson's son Galen puts the objection well:

> Although our thoughts about determinism appear in actual fact quite impotent to disturb our natural and unconsidered reactive attitudes and feelings (this reveals one commitment), it also seems very difficult for us not to acknowledge that the truth of determinism... bring the propriety of the reactive attitudes seriously into doubt (this reveals the other commitment). Defenders of the reactive attitudes may be unwise to seek to strengthen their position by

[21] It is not clear that it is correct. Pereboom for instance, thinks that whilst the realisation that we do not have free will would lead to some revisions of our attitudes, we could still sustain meaningful relationships for there are analogues of many of our reactive attitudes that do not presuppose the falsity of determinism (2001, ch. 4; 2002).

[22] Powys suggests that the art of happiness for instance, requires that we turn away from the truth:

> Are we to bow down meekly before this Moloch, crying, "blessed is the bitter truth! Let us and our dreams pass and perish, so long as the holy, horrible truth is recognised for what it is!" A thousand times no! If *this* is the truth of things, let us hug fiercely and obstinately each his own life-illusion! (1974, p. 17 and quoted in Smilansky 2000, p. 266)

> appealing to the fact that commitment to the reactive attitudes is, unlike the opposed commitment, *practically* basic. For the incompatibilist... may then reply that, while the commitment they are concerned to stress is of an essentially more theoretical character, it appears to represent a simple *truth*. There is a very real conflict of commitment. (G. Strawson 1986, p. 89)

So, even if Strawson is right and it would not be rational to give up the reactive attitudes upon the news that determinism is true, that does not show the incompatibilist to be wrong – it does not show us that moral responsibility and free will do *not* require genuine alternative possibilities.

We can also note something else about Strawson's view. By making our *being* responsible depend upon our being held responsible, Strawson's view makes our responsibility dependent upon matters *external* to the agent. For it is not in virtue of some fact about the agent that they are blameworthy, but rather due to some fact about the tendencies of those who blame us. In this respect Strawson's view seems fundamentally misguided in exactly the same way as the consequentialist compatibilist was. As Richard Double puts it:

> If it is not enough to be told that the blame we receive is justified because it contributes to the greater good of our society (thereby raising the scapegoating objection to consequentialism), then it is not enough to be told that blame reveals a deep tendency of blamers (or even those blamed themselves) to feel and express their reactive attitudes. In both cases, whether we talk about justification or not, we try to account for blame by citing facts that are extrinsic to the person blamed. (Double 1996, p. 67)

Fischer, who otherwise describes his approach to the question of free will and moral responsibility as 'Strawsonian' is sensitive to this kind of objection and himself criticises Strawson on the grounds that there need to be propriety conditions governing when it is and is not appropriate to consider someone blameworthy. It cannot be left to depend upon whatever our received practices are, leaving us without any tools with which to criticise or revise those practices.

> After all, once the actual application of the reactive attitudes is taken to be constitutive of moral responsibility, one wonders what should be said about situations in which communities hold people responsible who intuitively are not. Does the mere fact that certain attitudes are taken toward an agent establish that he is an appropriate candidate for this treatment? Imagine, for example, a society in which severely retarded or mentally disturbed individuals are resented, blamed, and harshly punished for their failure to adhere to the norms of the community... The problem here is that... there seems to be a difference between being *held* responsible and actually *being* responsible. (Fischer 1994, pp. 212-213)

Fischer goes on to propose 'proprietary conditions' on when it is, and when it is not, appropriate to blame and praise someone for something that they have done. The problem however, is that the instant one allows that propriety conditions are needed, then it is open to the incompatibilist to insist that genuinely available alternative possibilities are one of those conditions. In other words, there is nothing stopping an incompatibilist claiming that they too are 'Strawsonian' in Fischer's sense, but just disagreeing over the propriety conditions governing the appropriateness of the reactive attitudes.

In this section I have argued that P. Strawson's "reactive naturalism" should be rejected if proposed as the *whole* story about responsibility. For it too makes our responsibility too dependent upon external grounds, namely the fact that blaming and praising "reveals a deep tendency of blamers" (Double 1996, p. 67). To be acceptable Strawson's account needs to be supplemented with is some kind of proprietary rules for when it is and is not appropriate for someone to be blamed. However, the instant one concedes that such rules are needed, we are back in the old debate – for having genuine alternative possibilities and an ability to do otherwise could be maintained to be amongst the proprietary rules.

2.8 Reason Responsive Control

If we are rejecting Strawson's reactive naturalism, at least as the *whole* story about moral responsibility then the compatibilist seems committed to either rejecting one of the common sense principles contained in the consequence argument, or having to give their conditional analyses of 'could have done otherwise'.

However, in recent years many contemporary compatibilists have taken an altogether bolder approach to the question of control and alternative possibilities. Rather than pursuing either of the strategies above, they have instead *rejected* that moral responsibility requires alternative possibilities in any sense. Instead, all moral responsibility requires is that an agent's decision be in some way appropriately connected to their prior reasons. If connected in the right way – so if the process leading to decision is of the right kind - then the agent can be said to have controlled their decision, and can be deemed responsible for having made it.

I will elucidate the notion of control shortly. For one's initial reaction is likely to be to object that the compatibilist cannot just *reject* the requirement for alternative possibilities: if one's position is to be plausible one needs an argument showing that alternative possibilities are irrelevant. Given just how deeply rooted the association between alternative possibilities, free will, and moral responsibility is, the argument had better be a very good one.

But this is exactly what this modern breed of compatibilists believe themselves to have. The argument in question is one that has come to dominate contemporary discussions on free will. It involves a certain sort of case, known as a Frankfurt-style case, so called after Harry Frankfurt who first conceived them.[23] These cases are supposed to constitute counterexamples to the principle of alternative possibilities. I am going to discuss these cases in detail in the next chapter, but it would be as well to let Frankfurt present outline his case here:

---

[23] Actually Frankfurt's example is a variation of a case first presented by John Locke:

> 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…
>
> 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. (Frankfurt 1969, pp. 835-36)

So, in a Frankfurt-style case there is some person, Black, who does not intervene in a process leading to an agent's action, but would have intervened if the agent had been about to decide to perform some different action. Given that the counterfactual intervener plays no role in the agent's deliberations and subsequent action, it seems intuitive to consider that the agent is morally responsible for their action. The mere presence of the intervener seems, intuitively, to be irrelevant, yet because of their presence it also seems true to say that the agent could not have done otherwise. So, we have a case in which it is intuitive to consider someone blameworthy, despite the fact that they could not have done otherwise.

These cases appear to refute the principle of alternative possibilities. We do not seem to need the open pathways into the future – one pathway will do. If this is true, then the major motivation for thinking determinism rules out moral responsibility has gone, and much of the discussion of the problem of free will has been misguided. We might still *want* to have alternative possibilities; we might think that in the absence of genuine alternative possibilities we lose a valuable kind of free will, but the fact remains – or so

---

[A] man be carried whilst fast asleep into a room where is a person he longs to see and speak with, and be there locked fast in, beyond his power to get out; he awakes and is glad to find

argues this particular brand of compatibilist – that one still has the kind of free will and control presupposed by responsibility (Fischer 1994, p. 214).

With regard to the question of luck, such cases imply that ultimate luck is not a problem. For the only way to prevent ultimate luck is if one is armed with a genuine ability to do otherwise. But Frankfurt-style cases show such an ability to be irrelevant to the question of responsibility.[24] Thus likewise, exposure to ultimate luck is shown to be irrelevant.

I am going to be critically assessing Frankfurt-style cases in the next chapter, therefore I will restrict myself to giving a flavour of the kind of control that the Frankfurtian compatibilist claims is all that is needed for moral responsibility.

The first thing we can say, is that the modern compatibilist takes the important lesson from Frankfurt-style cases to be that one should focus on what happens in the actual sequence. Namely, "what the agents actually do, and how their actions come to be performed" (Fischer and Ravizza 1998, p. 37). Now some, known as 'mesh' theorists, hold that what matters is that a certain internal harmony (mesh) existed between, say, the agent's first and second order desires, or between their decision and values. I am going to say something more about those who go down this route in Chapter 4, so I will say nothing more here. Instead what I am going to do is provide only the very roughest of sketches of a reason-responsive model of control. Others have more fully developed such accounts.[25] My object is just that we should get some flavour of the model of control in question. For throughout the rest of the thesis, when I talk about compatibilist control, I will mean some kind of reason responsive control.

---

himself in so desirable company, which he stays willingly in. (1965, Book 2, ch. 11 sec. 10)

[24] Not all accept this. There are some – sometimes referred to as *hyper* incompatibilists – who agree that genuine alternative possibilities are not needed and that Frankfurt-style cases demonstrate this (for instance Hunt 2000; Stump 1999; Pereboom 2000). However, they argue that we still need to initiate our actions if we are to be responsible, and if determinism is true then we do not originate our acts. They may in addition claim that it is through originating our actions that we exercise a special 'luck eradicating' species of control over them. I find it hard to assess this latter claim for two reasons that will be elaborated upon later in the thesis. Firstly, I am sceptical about the coherence of these 'special' forms of control (I am a compatibilist about control). These concerns are aired in Chapter 5. Secondly I do not believe that it is possible that one could originate one's action *and* one not have genuinely available alternative possibilities. Frankfurt-style cases purport to present us with situations in which this is the case – but I believe that they fail to do this, an argument I will be making in the next chapter.

Take two darts players, Alan and Michael. Alan is an excellent shot, whereas Michael is useless. Alan nearly always hits what he is aiming at whereas Michael nearly always misses. In their respective actual sequences, both Alan and Michael take aim at the bulls eye, and both hit it. However, we would want to say that Alan's strike was far more controlled than Michael's. The reason here cannot be to do with any ability to do otherwise possessed in the actual sequence – for we can imagine that both Alan and Michael had a counterfactual intervener monitoring their aiming process, ready to intervene should either show any indication that they will miss (alternatively, one could just imagine determinism to be the case). Rather, the reason Alan's throw was more controlled was because it issued from a far more sensitive or discriminating mechanism or process. We ascertain this by asking, for instance, would Alan have still thrown his dart in the way that he did if the dart board was, say, a few millimetres further to the left? Alan, we can imagine, would have thrown his dart a few millimetres to the left in this case, whereas Michael would not (and so would have missed the bulls eye in this counterfactual sequence). Now what we are doing here is engaging once again in a counterfactual analysis. But this time we are employing such an analysis just to gauge what kind of mechanism we have operating in the actual sequence, namely how responsive or sensitive it is. In other words, we are just trying to get a handle on what it means to say that Alan's throw was more controlled than Michael's. We can say, on the basis of our counterfactual analysis, that when it comes to aiming Alan is far more sensitive to changes in his circumstances than ever Michael is.

We can translate the above to the case of an agent's decision making process. Let us call the 'mechanism' of an agent's decision, the process that brings it about (Fischer and Ravizza 1998, p. 38n). Following Haji (1998, p. 75) and Mele (1995, p. 177) we can characterise intentional deliberative action as having the following constituents: A psychological basis for evaluative reasoning, like desires, wants, and beliefs. These are psychological attitudes of the agent that can be cited as a reason why they made the decision that they did (he decided to make a cup of tea because he wanted one).[26] Then

---

[25] Those who hold or who have developed such accounts include: Dennett 1984; Fischer and Ravizza 1998; Haji 1998; Finagrette 1972; Gert and Duggan 1979; Neely 1974.

[26] Here we can distinguish between having a reason in an internalist sense and having a reason in an externalist sense. I may have a reason to make myself a cup of tea in the externalist sense if having a cup of tea (as some medical reports show) reduces my risk of developing cancer, even though I am completely unaware of this fact. I have an 'internalist' reason when I actually have a particular belief or desire. This may not actually move me to action, but I *have* it, in the sense that it could potentially have motivated me. In the debate over free will it is internalist reasons that are the principle focus, as it is only internalist

there is the judgement about which course of action to pursue, and a decision made on the basis of such a judgement. A decision can be said to *terminate* a deliberative process. The evaluative reasoning itself can be said to be a function of the agent's 'background evaluative scheme' which Haji defines as being composed of four constituents:

> 1) beliefs about normative standards to be invoked in assessing reasons for action; 2) motivational factors like long-term ends or goals that comprise the agent's values; 3) beliefs about deliberative principles the agent regards as appropriate to arrive at practical judgements about what to do or how to act in particular circumstances; and finally 4) motivation to act on the basis of the normative standards in 1 and values in 2 using the deliberative principles in 3. (1998, p. 125)

The agent's background evaluative scheme is what is employed to assess their desires and so forth and arrive at a judgement about what to do, or how to act.

The idea is that an agent's decision is controlled if it is appropriately sensitive to the agent's practical deliberations. To assess whether the mechanism leading to decision in a particular sequence is one that is appropriately sensitive one holds fixed the motivational precursors of the decision (so the agent's desire) and one holds fixed the agent's background evaluative scheme, and one considers whether there are scenarios in which, keeping these features fixed, the agent decides differently. If there are then the agent's decision in the actual sequence can be said to be responsive to the agent's reasoning process. Whereas if there are no scenarios in which the agent, possessing the same motivational precursor, acts otherwise, then the agent acts on the basis of an irresistible desire, and as such did not exercise control in the actual sequence (see Haji 1998, pp. 75-85).

It would be as well at this point to distinguish between moderate and strong reason responsiveness. A mechanism can be said to be strongly reason responsive where it always tracks, say, moral reasons. If, then, there is a moral reason not to X, then this

---

reasons that can actually feature in an explanation of the agent's action (we cannot explain why I made myself a cup of tea by citing the externalist reason that drinking such tea reduces my risk of developing

mechanism will always issue in the decision not to X. Now, in one sense this decision is perfectly controlled – for as mentioned above what the reason-responsive account gauges is the sensitivity of the agent's decisions to their reason. However, strong reason responsiveness is clearly an implausibly demanding control requirement and it would have the upshot that if an agent has a moral reason to do X, yet fails to do X, then they are not responsible. In other words, if strong reason responsiveness were the control requirement for moral responsibility then only the perfectly virtuous could be responsible, and it would be impossible to be responsible for wrongdoing.[27] As Fischer and Ravizza put it, a plausible control requirement would only be for a degree of reason responsiveness that allows for a "looser kind of fit between reasons and action" (1998. p. 43).

As it stands the account stands in need of various refinements if it is to deliver the intuitively correct results in a sufficiently broad range of cases.[28] As I say, this is a task that Fischer, Ravizza and Haji amongst others, have undertaken. But I believe we can and should accept that the above is an articulation of a type of control, a type of control which we can have under determinism. We should accept that, whether determinism is true or not, we do exercise this type of control over what we do. We are, in short, reason-responsive and thereby exhibit control over our decisions. Furthermore, we should also accept that reason-responsiveness of some degree is clearly a *necessary* requirement of being a responsible agent. The important question is over the sufficiency of this control. For nothing about this form of control shows how our possessing it does anything to protect us against ultimate luck. Reason-responsive control of the kind

---

cancer).

[27] By perfectly virtuous I do not mean to rule out internal conflicts. An agent's mechanism can be said to be strongly reason responsive even if the agent feels the tug of opposing reasons and has to make an effort to do the right thing. What makes for strong reason responsiveness is that the mechanism will issue in the right decision despite internal conflicts. This means that strong reason responsiveness might not involve perfect virtuousness by some virtue-ethicist lights.

[28] There are a number of difficulties for the account. Certain 'agents' need to be ruled out, and reason-responsiveness by itself does not seem to be able to do this. For example, there are some people who may be very *oddly* reason-responsive. I might do otherwise only if I am offered a £100 incentive, no more, no less. It is not at all clear whether a mechanism that is responsive in *that* way can be considered to be of the right kind. As Fischer and Ravizza put it:

> In judging a mechanism's receptivity [to reasons], we are not only concerned to see that a person acting on that mechanism recognizes a sufficient reason in one instance; we also want to see that the person exhibits an appropriate *pattern* of reasons recognition. In other words, we want to know if (when acting on the actual mechanism) he recognises how reasons fit together, sees why one reason is stronger than another, and understands how the acceptance of one reason as sufficient implies that a stronger reason must also be sufficient. (1998, pp. 70-71)

outlined is clearly not a form of *ultimate* control. As I mentioned above, Frankfurt-style cases seem to support the claim that we do not *need* such ultimate control in order to be in possession of the kind of free will relevant to moral responsibility. But much depends upon whether Frankfurt-style cases actually show what they purport to show. That is a matter that I will take up in the next chapter.

2.9 Conclusion

I have made the argument that determinism rules out alternative possibilities and considered various compatibilist responses. I dismissed those compatibilists whose strategy is to deny one of the intuitive principles employed in the Consequence Argument, and I raised some standard objections to some of the other proposals, before finally going on to outline the kind of reason-responsive control that the most recent breed of compatibilists hold is sufficient (in terms of control) for free will and moral responsibility. Critical assessment of the Frankfurt-style cases which inspired their positions will occupy the next chapter.

Chapter 3

Frankfurt-Style Cases

In the previous chapter we saw there is a new breed of compatibilist who hold that a certain kind of case – the Frankfurt-style case – demonstrates the irrelevance of alternative possibilities to the question of moral responsibility. On the basis of such cases these compatibilists advocate the abandonment of the 'garden of forking paths' view of free will, at least where moral responsibility is concerned.

However, a lot depends upon whether Frankfurt-style cases actually work. In what follows I argue that the original Frankfurt-style case does *not* work. I will consider a number of recent attempts to improve upon the original, and I will argue that all but one fail too.

Fischer's variation *might* work. But if it does, it has limited power. Fischer's case suggests at most that responsibility does not require path-picking control. It does *not* challenge the idea that responsibility requires genuine alternative possibilities. In other words, Fischer's case only *half* works, if it works at all.

The intuitive association between alternative possibilities, free will, and moral responsibility is a reflection of two demands. First a demand for independence or origination. Second a demand for path-picking control. If the arguments that follow are right, Fischer's case does not challenge the need for alternative possibilities, but rather why we need them. It may be more to do with the first demand, rather than the second.

3.1 The Case

Although I gave Frankfurt's original case in the previous chapter, it would be as well to provide it again here:

> 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...
>
> 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. (Frankfurt 1969, pp. 835-36)

To reiterate, what we have in the above situation, and what is characteristic of all Frankfurt-style cases, is the presence of some counterfactual intervention device – be it a person, a drug, phobia, or post-hypnotic suggestion. In the actual sequence this device plays absolutely no role in what happens: it is *as if* it is not there. For that reason its mere presence seems irrelevant to any assessment we might make of the agent's blameworthiness. However, its mere presence *does* mean that the agent could not have decided otherwise, and thus PAP appears to have been refuted.[29]

I will first consider some inadequate defences of PAP, which nevertheless prompted some minor adjustments to Frankfurt's original case. I will then go on to consider what I take to be an altogether more devastating criticism before considering a number of recent responses to it.

---

[29] Frankfurt – a compatibilist – holds that what his examples highlight is the PAP needs to 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. 839). But this revised principle is *also* refutable by Frankfurt-style cases, if, that is, they refute PAP. For too often it is assumed that it is in some way essential to a Frankfurt-style case that the agent does not know about the counterfactual intervener (in fact, to my knowledge *everyone* has assumed this). But this is actually not essential. Imagine a Frankfurt-style case in which Jones *knows* that Black is going to make him decide to do X, unless he does X under his own steam. Jones decides to X under his own steam, precisely because he believed that Black was going to make him decide to X anyway. But if X is a morally reprehensible decision then isn't Jones morally responsible? I certainly have the intuition that he is. Yet Frankfurt's principle is violated in the case as I have described it. For Jones only decided as he did *because* he thought it was going to be inevitable that he would make the decision in question. (I make this point in my 2004, though I develop it there in a way that I would now no longer endorse.)

3.2 Some Inadequate Defences of PAP

The most straightforward way to undermine the power of an Frankfurt-style case is to claim the situation involved does actually allow for the possibility of the agent doing otherwise. With this in mind, the first such defence of PAP that I want to look at here begins with an assumption about what the agent is responsible *for*, and assumes that the agent is responsible for the occurrence of some particular event. This defence involves assuming that some or all of the causal antecedents of an event are essential to the individuation of an event. On this view if an event has different causal antecedents (or relevant different causal antecedents) from an otherwise identical event then it is not the same event. This is a controversial view about event individuation, but there is no reason to think that a PAP defender would not mind being committed to it, if that is what it takes to defend PAP. With these assumptions in place we can return to the Frankfurt-style case and notice that in the actual sequence the agent causes some event to occur (say, the assassination of the president) by themselves. However, in the alternative sequence in which the counterfactual intervener makes the agent assassinate the president, the event of the president being assassinated is different for it has a different causal history. The argument now is that the agent does have the power to cause some different particular event to occur. It is true that the agent does not have the power to prevent himself from assassinating the president, however he does have the power to bring about some alternative event (van Inwagen 1983, pp. 166-171 and see Fischer 1994 p 137).

The alternative possibility in question is not one that a defender of Frankfurt-style cases would necessarily reject (though the particular account of event individuation it assumes could be questioned on independent grounds (see Carter 1979 pp. 443-452). It is rather, as Fischer says "that this alternative possibility is not sufficiently robust to ground the relevant attributions of moral responsibility" (1994, p. 140). Consider the following, again from Fischer:

> The existence of various genuinely open pathways [into the future] is alleged to be crucial to the idea that one has control of the relevant kind. But if this is so, I suggest that it would be very puzzling and unnatural to suppose that it is the

> existence of various alternative pathways along which one does not act freely that shows that one has control of the kind in question. How exactly could the existence of various alternative pathways along which the agent does not act freely render it true that the agent has the relevant kind of control ?... [N]ote that even if it is granted that the terminus of the alternative sequence in the case of Jones and Black is a different event from the actual event of Jones' voting for Clinton, it also is evident that Jones would not be freely voting for Clinton in the alternative sequence. (1994, p. 141)

The point is that even if the existence of such an alternative possibility is granted, it does not seem capable of doing the work required of it, namely grounding the agent's moral responsibility. For this particular flicker strategy "confuses the ability deliberately to do otherwise with the possibility of something different occurring" (Fischer 1982).[30]

There is an important point here. We do, when we conceive of ourselves as free, take ourselves to have alternative possibilities in which we *freely decide* otherwise. In other words, we take ourselves to have alternative possibilities of a certain robustness. As such it is not enough for a PAP defender to alight on just any alternative possibility and insist that it shows the necessity of the PAP condition. For if we can never freely decide otherwise, because, for instance, a counterfactual intervener is waiting in the wings for all of our decisions (a situation referred to as a 'global' Frankfurt-style case), then this is, in every way worth caring about, to be as big a cramp on our freedom as determinism would be.

However, perhaps the real point being made by the PAP defender is that even if the agent does lack the ability to freely decide otherwise, they nevertheless do have the ability to initiate an alternative action. For instance, if it is the case that, should the agent show any inclination of choosing otherwise, intervention will take place, the claim is that the agent does at least "initiate, (albeit not complete) the choice to do otherwise"

---

[30] It is also worth pointing out that it is irrelevant whether or not the counterfactual intervention involves forcing the agent to make a particular decision. It is true that in the original Frankfurt-style case Jones freely decides to do what Black would otherwise have forced him to decide to do. But it is irrelevant what Black will force Jones to decide to do in the counterfactual sequence, the case would be a counterexample to PAP even if Blacks' intervention took the form of killing Jones. The above criticisms of Frankfurt-style cases therefore seem to be focusing on what is an irrelevant feature of the standard Frankfurt-style case.

(Fischer 1994 p. 136). The ability to initiate a choice to do otherwise is arguably a robust enough alternative possibility to ground moral responsibility.

The problem though is that Frankfurt style cases can be reconstructed in a way that removes even this kind of possibility. For rather than it being some alternative initiating action on the part of the agent that triggers the intervention, it could be some involuntary sign, such as a blush or twitch (Blumenfeld 1971 p. 340-1). Here is Fischer's version of a 'prior sign' Frankfurt-style case:

> How can the [counterfactual intervention] device tell whether Jones is about to choose to vote Republican or Democrat? This is where the "prior sign" comes in. If Jones is about to choose at t2 to vote for Gore at t3, he shows some involuntary sign – say a neurological pattern in his brain – at t1. Detecting this, Black's device does not intervene. But if Jones is about to choose at t2 to vote for Bush at t3, he shows an involuntary sign – a different neurological pattern in his brain – at t1. This brain pattern would trigger Black's device to intervene and cause Jones to choose at t2 to vote for Gore, and to vote for Gore at t3. (2003, p. 192)

This 'sign' is not an action – it is not something the agent chooses to display. Therefore, if we reconstruct the Frankfurt-style case using a prior sign as the trigger rather than an 'initiating action' the flicker of freedom in question disappears. For the prior sign is not voluntarily displayed and therefore it is not something that it is within the power of the agent to display.

However, whilst these early criticisms of Frankfurt's original case did little more than prompt clarification and minor adjustment, another criticism poses far more serious problems. These are problems to which I believe Frankfurt-style cases, in their normal form, cannot overcome.[31]

---

[31] One line of response that I have not considered here maintains that PAP is too deeply entrenched to be rejected on the grounds of our intuitions about Frankfurt-style cases (Copp 1997). This argument usually draws support from the association between PAP and another, equally well entrenched principle – the Kantian principle that 'ought' implies 'can'. For after all, if ought implies can, then ought not implies can refrain. But if we do sometimes do wrong, then this means that it must have been possible for us to refrain, and hence the need for alternative possibilities is generated (see Widerker 1991; Haji 1998 & 2002; Schnall 2001; Fischer 2003a). The point then is that because Frankfurt-style cases threaten PAP, they also threaten the Kantian principle, and rejecting both principles is just too costly, so we should reject our

3.3 The Kane/Widerker Defence of PAP

It has now become "increasingly common" to reject Frankfurt-style cases on the basis of the following argument (Warfield 1996, p. 221). The argument has been made in one form or another by Kane (1985 and 1996), Widerker (1995a and 1995b), Ginet (1996), and Wyma (1997). For the sake of simplicity I will concentrate on Kane and Widerker's versions of the argument.

The first step in this argument is to recognise that if a Frankfurt-style case is to have any bite against an incompatibilist defender of PAP it cannot assume the truth of determinism. The whole point of a Frankfurt-style case is that it presents us with a scenario in which an agent is morally responsible irrespective of the presence of a counterfactual intervener. That means that, absent the counterfactual intervener, the situation needs to be one in which PAP is met for otherwise the case begs the question against the incompatibilist by just assuming that moral responsibility does not require alternative possibilities. In other words, the situation needs to be one in which, absent the counterfactual intervener, the agent *genuinely* could have decided otherwise at the moment of choice, holding fixed all prior states and the laws of nature. Bearing this in mind, Kane makes the following observation:

> The controller, Black, plans to make Jones do A. But he waits to see if Jones is going to do A on his own and only intervenes if Jones is about to do B instead... [But] the controller cannot know which one is going to occur beforehand unless he predetermines one of them to occur. He can therefore wait until he finds out whether [Jones] will do A or B, but then it is too late to

---

intuitions instead. However, the problem with this line of response is that those who reject PAP might be equally inclined to reject the Kantian principle too. After all, the intuitive appeal of the Kantian principle surely lies in the intuitive appeal of PAP, such that if one rejects one, one is likely to reject the other (see Fischer 1999a and 2003a for this kind of response). The Kantian principle can, after all, be criticised on independent grounds, such as the possibility of genuine moral dilemmas (see Sinnott-Armstrong 1988). Added to which we can note that control demands for moral responsibility are in general more stringent than those for 'deontic' morality (judgements of right, wrong, ought and ought not). The possibility of excuses testifies to this (if one has an excuse one has done wrong, but is not blameworthy for having done so). So again, those who reject PAP for responsibility, are unlikely to think PAP *is* a requirement of deontic morality (though see Haji who holds exactly this position and develops it in his 2002). Some have also argued that it is possible to detach the Kantian principle from PAP, and in this way preserve the Kantian principle whilst rejecting PAP (see Frankfurt 1988, p. 98 and Yaffe 1999). I think such attempts

> control the choice. Or he can intervene in the brain, shutting down the indeterminacy or its effects before either choice occurs, thereby determining the outcome he wants. In the latter case the choice will be determined by the controller and the controller, not the agent, will be *ultimately responsible* for it... By contrast, if the controller does not intervene to predetermine the outcome and the indeterminacy remains in place until the choice is made... then the agent, and not the controller, is ultimately responsible for it. But then it is also the case that the agent *could have done otherwise*. (1996, p. 142)

In other words, if the agent's decision making process is internally indeterministic up to the moment of choice, then the agent will have alternative possibilities at the moment of choice, despite the presence of the intervener.[32]

Widerker concentrates on the exhibition of a prior sign to make the same point (1995a; 1995b). The prior sign is supposed to provide Black with a basis for deciding whether to intervene. If Black knows that a twitch from Jones at T1 entails that he will perform the action that Black wants him to perform at T2 (and that therefore intervention is unnecessary) then the relationship between the sign and the subsequent decision made by Jones must be deterministic. The sign must either deterministically cause Jones to perform the action in question "or is in some way associated with something at t1 that deterministically causes [Jones to perform the action]" (Mele and Robb 1998, p.100). But if this is so the question has been begged. If, however, the exhibition of the prior sign does not in any way determine Jones's subsequent decision "it is hard to see how Jones's decision is unavoidable" (Widerker 1995a, p. 251). If the Frankfurt-style case is not to beg the question it must be the case that it is possible for Jones to exhibit the prior sign (indicating that he is going to do what Black wants him to do such that Black does not intervene) and Jones decides not to perform the action in question: he does otherwise (Widerker 1995a, pp. 250-1).

---

doomed, but again, I will not make such arguments here (Fischer responds to Yaffe in his 2003, as does Haji in his 2002, pp. 54-58).

[32] One response from a Frankfurtian (though not one that has been explicitly made) might be to accept Kane's point, but to then go on to point out that *this* is what the incompatibilist is driven to in order to protect PAP from refutation. The incompatibilist has to insist upon indeterminism right up to the moment of choice, and this, the Frankfurtian might claim, is going to give the incompatibilist serious problems when it comes to explaining exactly how the agent can have controlled the resulting decision rather than it just being random. In other words, Frankfurt-style cases force the incompatibilist into an embarrassing situation. The concern about indeterminism and control is taken up in Chapters 5 and 6.

What makes this argument so strong is that in the alternative scenario in which Jones does do otherwise and does exhibit the prior sign, Jones is acting freely. The alternative possibility that Kane, Widerker and others have uncovered is therefore of the most robust kind.

I believe that this kind of argument is devastating to Frankfurt-style cases. However, Frankfurtians are a committed band of Frankfurt-style case-users and between them they can muster considerable ingenuity in defence of their counterexamples. It would be foolhardy to think that the above argument would be left as the last word on the matter. Since the question begging accusation has been made several responses have been forthcoming. I will critically assess four of the most prominent 'rescue' attempts. The first three involve the presentation of a new, or more detailed variation of a Frankfurt-style case in which agent-internal indeterminism is preserved without this jeopardising the counterfactual intervener's ability to rule out alternative possibilities. In other words, they purport to present us with a case in which the agent's deliberative process is indeterministic, yet it is also the case that the agent could not have made a different decision from the one that they did make. The final kind of case I will consider is altogether bolder, insofar as determinism is explicitly assumed and the question begging charge is challenged head-on.

3.4 Mele and Robb's Rescue Attempt

Let me start by considering the prominent 'rescued' Frankfurt-style case developed by Mele and Robb. In Mele and Robb's ingenious variation the agent Jones's decision making process is indeterministic but, simultaneous with Jones's decision making process runs another, deterministic process. In the actual sequence, Jones's indeterministic process issues in Jones making a particular decision (in their particular case this is a decision to steal Ann's car). However, the deterministic process would have issued in the same result had the indeterministic process not. We can let Mele and Robb describe exactly how this could be:

> At T1, Black initiates a certain deterministic process P in Jones's brain with the intention of thereby causing [Jones] to decide at T2 (an hour later, say) to steal

> Ann's car. The process, which is screened off from Jones's consciousness, will deterministically culminate in [Jones's] deciding at t2 to steal Ann's car unless he decides on his own at T2 to steal it or is incapable at t2 of making a decision (because, e.g., he is dead by t2). (Black is unaware that it is open to [Jones] to decide on his own at T2 to steal the car; he is confident that P will cause [Jones] to decide as he wants [Jones] to decide.) the process is in no way sensitive to any 'sign' or what [Jones] will decide. As it happens, at T2 [Jones] decides on his own to steal the car, on the basis of his own indeterministic deliberation about whether to steal it, and his decision has no deterministic cause. But if he had not just then decided on his own to steal it, P would have deterministically issued, at T2, in his deciding to steal it. Rest assured that P in no way influences the indeterministic decision-making process that actually issues in [Jones's] decision. (1998, pp. 101-2 & see also their 2003)

Intriguing though this variation of a Frankfurt-style case is, I do not believe it can work without assuming determinism at a critical juncture. Consider first, that it is open to the incompatibilist to maintain that the agent needs to have been able to decide otherwise *up to the moment of decision-making*. So, in other words, if one were to rewind the clock to the moment just prior to decision-making and then rerun the sequence again, it would be possible for the agent to decide otherwise. As far as I can see Mele and Robb's case does beg the question against this kind of incompatibilist, for Mele and Robb's case has to assume a deterministic relation prior to the decision-making moment. To see this, we can turn to an illustration given by Mele and Robb.

Mele and Robb imagine that someone might wonder how Jones can decide on his own in the actual sequence, without the deterministic process having caused him to decide as he did. They provide the following answer.

> Consideration of the following fanciful machine will prove useful in answering this question. The machine, designed by a specialist in machine art, produces artistic widgets of different shapes and colors. The colors of the widgets produced are determined by the color of a ball bearing (bb) that hits the machine's receptor at a relevant time. The machine, M, is surrounded by several automatic bb guns, each containing bbs of various colors. The relevant aspect of

> M's mechanical design, for our purposes, is relatively simple. First, with one qualification, if a bb of color x hits M's receptor, and M is not already in the process of making a widget, M t once starts a process designed to result in the production of an x-colored widget. Second, because two or more bbs sometimes hit the receptor simultaneously, the artist has designed his machine in such a way that whenever this happens (while M is not busy making a widget) M at once starts a process designed to result in the production of a widget the color of the right-most bb... [Jones] is analogous to M in an important respect. He is physically and psychologically so constituted that if an unconscious deterministic process in his brain and an indeterministic decision-making process of his were to "coincide" at the moment of decision, he would indeterministically decide on his own and the deterministic process would have no effect on his decision. This situation is an analogue of a case in which two bbs of the *same color* simultaneously hit M's receptor. (1998, pp. 103-104)

There is a sleight of hand in the above example. Strictly speaking, in the mechanical example we have two bbs hitting M's receptor simultaneously, and this prior event then results in M issuing a blue widget. The blue widget is the analogue of Jones's decision. So in fact we do not have two processes 'coinciding at the moment of decision', rather we have two processes coinciding *just prior* to decision making. This is crucially significant. If this prior event – the event analogous to two bbs striking the receptor at the same time - determines that Jones will make a particular decision, then irrespective of whether the sequence of events leading up to this point was indeterministic or deterministic, Jones is not morally responsible for this decision. For note, on the incompatibilist view in question, Jones needs to have the genuinely available alternative possibility of deciding otherwise at the moment of decision.[33] If his own deliberative process is one that becomes deterministic prior to decision-making, then Jones will lack

[33] This need not always be the case, but if an agent is to be responsible for what they do, then their action must ultimately trace to acts which were indeterministic at the moment of choice. To avoid always having to put in this mouthful of words I will take it for granted that Frankfurt-style cases need to work for these actions, so we can assume that it *does* have to be indeterministic at the moment of choice or else the question will have been begged against most incompatibilist positions. It is possible to defend a variation of incompatibilist position according to which it *doesn't* need to be genuinely possible for the agent to have decided otherwise at the moment of choice. We will see later that Mele has proposed such a variation. I think there are serious problems with such positions, not least of which would be their susceptibility to Frankfurt-style cases. Part of the value of Frankfurt-style cases can be seen to lie in their forcing the incompatibilist to recognise that it needs to be indeterminate *up to the moment of choice* which decision the agent will make.

the required alternative possibilities. Thus, even in the sequence in which Jones makes the decision under his own steam, Jones will not be responsible. At the crucial moment, have lacked robust alternative possibilities: not due to the presence of the deterministic process, but due to the presence of determinism in his *own* deliberative process.

It could be that we can retain indeterminism at the moment of choice by (analogously) making it indeterministic what happens when a bb hits the right-most side of the receptor. In this case it may or may not issue in a blue widget. In other words, when Jones's own deliberative process reaches point A (analogous to a bb hitting the receptor), it is *indeterministic* whether this will result in Jones making decision X or decision Y. But if this is so, then Jones *does* have robust alternative possibilities in the sequence in which he makes a decision under his own steam. The presence of the deterministic process does nothing to alter this.

It seems to me that Mele and Robb's case does not improve upon the original Frankfurt-style case. Exactly the same dilemma arises. Either determinism has to be assumed at a crucial juncture (namely at some point prior to decision making) – in which case the question has been begged – or indeterminism is left in place, in which case Jones retains robust alternative possibilities at the moment of choice.

What Mele and Robb need to do, is show how it is possible for the deterministic process not to cause Jones to make the decision to steal Ann's car in the actual sequence, yet at the same time maintain that it would have caused Jones to make the decision to steal Ann's car had Jones been about to decide otherwise. I do not see how this can possibly be done. Mele and Robb claim that the deterministic process will not cause the decision, if Jones makes it under his own steam. But this requires a pre-decision making coincidence of the deterministic and indeterministic processes – and it is this which undoes their case. For they have to smuggle determinism into Jones's *own* decision making process. Let us turn to the second of the four revived Frankfurt-style cases, this time one devised by Eleanor Stump.

3.5 Stump's Case

Eleanor Stump's version of a Frankfurt-style case focuses on mental acts (1990; 1995; 1996; 1999). Mental acts are correlated to a complex series of neural firings. As Stump says "[w]hen I suddenly recognise my daughter's face across a crowded room, that one mental act of recognition, which feels sudden, even instantaneous, to me, is correlated with many neutral firings as information from the retina is sent through the optic nerve, relayed through the lateral geniculate nucleus of the thalamus, processed in various parts of the occipital cortex, which take account of figure, motion, orientation in space, and color, and then processed further in cortical association areas" (1999, p. 417). But the interruption of such a series means that there is no mental act at all, rather than there being 'part' of a mental act. She explains below:

> If the neural sequence correlated with my recognising my daughter's face across a crowded room is interrupted at the level of the thalamus, say, then I will have no mental at having to do with seeing her. I won't for example, think to myself, "For a moment there, I thought I saw my daughter, but now I'm not sure." I won't have a sensation of almost but not quite seeing her. I won't have a premonition that I was about to see her, and then I mysteriously just don't see her. I will simply have no mental act regarding recognition of her at all. (Stump 1999, pp. 417-418)

If we grant all this, then a seemingly non-question begging Frankfurt-style case can be constructed in which a counterfactual intervener can detect from the complex series of neural events what mental act it will be. As Fischer puts it on Stump's behalf:

> If... the counterfactually intervening liberal neurosurgeon, did interrupt a neural sequence that was beginning (and which is such that, if it were completed, it would constitute, or correlate with, a decision to vote for Bush), Jones would *not* (according to Stump) have engaged in the mental act of *beginning to make a decision.* Jones would have *no* mental act, just as Stump would not have begun to recognise her daughter, if the sequence of neural firings beginning in her retina had been terminated in the thalamus. (Fischer 2002, p. 295 and see also Stump 1999, p. 418)

But again, I fail to see how any of this guards against the question begging accusation. Presumably the point is supposed to be that the complex neural sequence which is identical to the mental act can itself have been indeterministically initiated. In this way it remains indeterministic up to the moment of decision which way the agent will decide. However, because the moment of decision actually turns out to be a prolonged neural sequence, this allows the counterfactual intervener space to work out what decision is about to be instantiated by this neural sequence. But then Stump, as with Mele and Robb, has smuggled in a deterministic relation at the crucial point. If this particular neutral process will deterministically result in Jones having made a particular decision, then at some point *prior* to the decision being made, Jones could not have decided otherwise. Hence, the counterfactual intervener is allowed their reliable basis for intervention only at the cost of making Jones's decision making process deterministic at a crucial juncture.

Perhaps Stump will reply that the neural sequence is not to be thought of as a process which culminates in Jones making a particular decision, but rather that the whole sequence – this neural sequence, that is – *is* the decision.[34] But if that is what Stump means, then Jones could have decided otherwise up to the moment of decision. Firstly, it would have to be the case that the neural sequence and the decision occur simultaneously – a point Widerker makes. For unless this is the case then the neural process is a deterministic cause of Jones's decision and the question will have been begged (see Widerker 2002, p. 326 fn. 8). Yet now it seems quite clear that the counterfactual intervener's basis for intervention comes too late. The intervener can only get to work so to speak, after the agent has exercised the relevant responsibility grounding power. Widerker makes the point below:

> Consider a scenario like the one described by Stump, except that it does not feature a counterfactual intervener... In that scenario, there would be no reason to think that Jones could not have decided otherwise or that he could not have refrained from the decision he made. Recall that, on Stump's alternative account of decisions, once the neural firings a, b, c occur, Jones is... bound to decide to

[34] Widerker makes this suggestion on Stump's behalf – only then to reject it on the same grounds as I am about to. See his 2002, p. 326.

> vote for a Republican candidate. This means that the only way in which Jones could have refrained from his decision in that scenario, is by having the power to bring about the non-occurrence of a, b, c, a power that he would have before the occurrence of a, b, c and not after that. But if he has that power in the said scenario (as he surely does), he must also have it in the scenario featuring [the intervener]. That the latter scenario includes a potentially coercive neuroscope dos not change this fact, since its coercive influence would come into play only... at a time later than the occurrence of a, b, c. Hence it does not affect Jones's power to bring about the non-occurrence of a, b, c. (2002, pp. 326-327)

As far as I can see Stump's Frankfurt-style case fails to do anything to meet the question begging charge. Just as with the original, and just as with Mele and Robb's variation, either determinism has to be assumed at a crucial point (a point just prior to decision making), or it has to be allowed that the agent has robust alternative possibilities in the actual sequence. So far PAP remains unthreatened.

3.6 Pereboom's Rescue Attempt

The third of the four rescue attempts I want to consider is also the final attempt at an 'indeterminism preserving' Frankfurt-style cases. It has recently been put forward by Derk Pereboom:

> Joe is considering whether to claim a tax deduction for the substantial local registration free that he paid when he bought a house. He knows that claiming the deduction is illegal, that he probably won't be caught, and that if he is, he can convincingly plead ignorance. Suppose he has a very powerful but not always overriding desire to advance his self-interest no matter what the cost to others, and no matter whether advancing his self-interest involves illegal activity. Furthermore, he is a libertarian free agent. But his psychology is such that the only way that in this situation he could choose not to engage in the tax evasion is for moral reasons. His psychology is not, for example, such that he could decide not to evade taxes for no reason or simply on a whim. In fact, it is causally necessary for his deciding not to evade taxes in this situation that a moral reason occur to him with a certain force. A moral reason can occur to him with that

> force either involuntarily or as a result of his voluntary activity (e.g. by his willing to consider it, or by his seeking out a vivid presentation of such a reason). But a moral reason occurring to him with such force is not causally sufficient for his deciding not to evade taxes. If a moral reason were to occur to him with that force, Joe could, with his libertarian free will, either chose to act on it or act against it (without the intervener's device in place). But to ensure that he decide to evade taxes, a neuroscientist now implants a device which, ere it to sense a moral reason occurring with the specified force, would electronically stimulate his brain so that he would decide to evade taxes. In actual fact, no moral reason occurs to him with such force, and he chooses to evade taxes while the device remains idle. (2001, quoted in Fischer 2002, pp. 297-298)

The idea here is that the occurrence of the moral reason of requisite strength only gives Joe the opportunity to exercise his libertarian free will by either acting on the basis of that reason, or not. So, if the reason occurs to Joe, the intervener knows that it is now *possible* that Joe will decide not to evade taxes (we can assume that the intervener wants Joe to evade taxes). Because the intervener does not want to take any risks, they intervene on this occasion *forcing* Joe to make the decision to evade. However, in the actual sequence the moral reason does not occur to Joe, and so there is no possibility of Joe deciding against evading taxes for moral reasons, and hence no need for intervention.

However, the first thing we can say is that if Joe's decision to evade taxes is, in the actual sequence, inevitable given Joe's internal situation (so, irrespective of the intervener's presence) then an incompatibilist is within their rights to hold PAP violated such that Joe is not morally responsible. The important question is not whether a Frankfurt-style case can be constructed which can rule out *some* alternative possibilities, but whether one can be constructed which rules out *relevant* alternative possibilities *without* having to assume that they are *already* absent. So, does Pereboom's case present any kind of counterexample to incompatibilists who maintain that it needs to be open, up to the moment of decision, which way an agent will decide? No. For either it is inevitable that Joe will make the decision that he does at the moment of decision making – in which case the question has been begged against at least some

incompatibilists, or it is not, in which case Joe does have the relevant kind of alternative possibilities available to him. [35]

Perhaps Pereboom will claim that indeterminism is preserved in the actual sequence up to the moment of decision making because there were alternative decisions that the agent could have made. It is just that in the actual sequence those alternatives would not have included deciding not to evade taxes for moral reasons – for the relevant moral reasons did not occur to Joe in the actual sequence. We can, in other words, understand the counterfactual intervener as one who is more concerned that Joe does not make a particular decision, rather than one who is concerned that Joe does make a particular decision.

Let us grant the above. In what sense, if the above is true, does Joe lack alternative possibilities? He does not. The intervener's presence only restricts Joe's option range, or rather, restricts the basis upon which he can make a decision. This kind of restriction can certainly have a bearing on an agent's responsibility. Moral responsibility is, after all, about more than free will and PAP is not a sufficient condition for responsibility. I would say that, irrespective of the presence of an intervener, if the moral reasons not to evade taxes do not occur to Joe, then Joe is not morally responsible for the decision that he makes in the actual sequence for Joe would not have had the possibility of acting for moral reasons.[36] He still acts with free will insofar as it was genuinely possible for him to have made a different decision at the moment of decision-making. [37] In the same way,

---

[35] We will see later that there are variations of incompatibilist position according to which the relevant indeterministic breaks in an agent's deliberative process (the process leading to decision making) can come earlier. For instance, it might be maintained that if it is indeterministic what beliefs 'come to mind' during an agent's deliberations then the agent has relevant alternative possibilities (a certain picture of the future as a future of branching pathways is secured this way). Pereboom's case does, I would hold, highlight the inadequacies of such a position – something I'll say more about later, for I believe there are other ways in which this kind of position can be shown to be inadequate. I know of no-one who explicitly defends such a variation of incompatibilist position (Mele outlines such a position, but does not endorse it).

[36] Perhaps one will claim that moral reasons *should* have occurred to Joe. But then, from the robust modest libertarian perspective (and the control-libertarian perspective also) this is claim that there was some earlier point at which Joe had the genuine possibility of acting for moral reasons, and had he done so moral reasons would have occurred to him at this point.

[37] One might object that in the actual sequence the agent could have had moral reasons occur to them through an act of their own will – Pereboom did, after all, say that Joe could seek out a "vivid presentation of such a [moral] reason". In other words, so long as Joe possesses libertarian control, it was possible for him to have brought such moral reasons into his deliberations, and as such it remains intuitive to consider him blameworthy in the actual sequence. But in response we can make two points. Firstly, this is a two-stage process: first Joe brings, through an act of will, the moral reasons into his deliberations, and *then* he has the opportunity to decide on the basis of such reasons. The counterfactual intervener will rule out the

someone who is completely insane might still act with free will, but lacks *moral* responsibility because their option ranges consist only of crazy options. So, in the case that Pereboom presents, the lack of a genuinely available alternative possibility in which Joe acts on moral reasons does, I would say, subvert Joe's responsibility. Pereboom has failed to present a case in which the agent is unproblematically morally responsible in the actual sequence. Pereboom's case provides us with no reason to think anything suspect about PAP.

3.7 Fischer's Rescue Attempt

Thus far the rescue attempts I have been considering have all tacitly accepted that determinism cannot be assumed in the construction of a Frankfurt-style case, and have sought ways in which to rule out alternative possibilities whilst leaving in-place agent-internal indeterminism.[38] But John Martin Fischer, (who has probably done more than anyone else to ensure Frankfurt-style cases remain prominent in the debate over free will) takes a different approach and challenges the question-begging charge (Fischer 1999a; 2002; forthcoming *a*).[39] Fischer claims that even if determinism is explicitly assumed in the construction of the Frankfurt-style case, this does not beg the question. To see his reasoning it is best to understand how he envisages the determinism-assuming Frankfurt-style case working.

To begin with, assume that determinism obtains. Imagine an agent making a morally reprehensible decision. An incompatibilist will not agree that the agent is morally responsible, due to the presence of determinism in the actual sequence. However, unless

---

second stage, and as such in the actual sequence the most that Joe can be blameworthy for is not having brought the moral reasons into his deliberations.

[38] One 'rescue' attempt that I have not considered in the main body of this chapter is Hunt's 'Blockage' case (the term was coined, I believe, by Fischer in his 1999a). In Hunt's case we assume that the agent makes a decision and that this employs certain neural pathways in the brain (we are assuming that the mind supervenes on the brain), but other pathways are blocked, and hence there was no possibility of the agent deciding otherwise (Hunt 2000). But on this one I follow Fischer when he says the following:

> [T]he example is difficult to imagine (and thus properly evaluate). If causal indeterminism obtains in the actual neural pathway, how exactly can it be the case that the agent does not have access to events consisting in the bumping up against any of the barriers (intermediate or terminal)? And if the agent really does not have access to any such 'bumping' events, how can it be the case that causal determinism does not actually obtain? (2002, p. 296 and see also 1999a, p. 119)

[39] Haji and McKenna have also recently endorsed such an approach, or at least, have argued that the assumption of determinism does not necessarily beg the question against the incompatibilist (see their 2004).

we are going to just start-out by assuming that free will and moral responsibility require the falsity of determinism then it would not be "dialectically kosher" to insist that the agent is not morally responsible (Fischer forthcoming *a*, and quoted in Haji and McKenna 2004, p. 309). Rather, judgement should be suspended.[40] With this suspension agreed to, we then have the counterfactual intervener added to the scene. We are invited to see that the mere presence of the intervener is irrelevant to any assessment we make of the agent's moral responsibility in this situation. In other words, even the incompatibilist can agree that the mere presence of the intervener is irrelevant to the agent's non-responsibility. This is all that is needed for the case to provide reason to abandon PAP, because both determinism and the presence of the intervener were sufficient to rule out alternative possibilities.

> [T]*wo* causes make it the case that Jones is unable to choose otherwise at T2: the prior condition of the world (together with the laws of nature) and Black's counterfactual intervention. What the examples show is that *the mere fact* that Jones is unable to choose otherwise does not *in itself* establish that Jones is not morally responsible for his choice. This is because Black's counterfactual intervention is one of the things that make it the case that Jones is unable to choose otherwise at T2, and yet it is irrelevant to the grounding of Jones's moral responsibility. Considering this factor (the counterfactual intervention), and bracketing any other factor that might make it the case that Jones is unable to choose otherwise at T2, it seems to me that Jones may well be morally responsible for his action. The mere fact that he lacks alternative possibilities cannot in itself be the reason that Jones is not morally responsible, if he is not morally responsible. (Fischer, forthcoming a, sec. 4.5)

Fischer capitalises on the fact that what really does the work of refuting PAP is the intuitive irrelevance of the counterfactual intervention machinery, rather than whether or not the agent is morally responsible in the actual sequence. He then points out that if alternative possibilities were the important ingredient needed for moral responsibility,

---

[40] Although both Haji, McKenna and Fischer seem to think this dialectical point is important, I fail to see why. Fischer's case works – if it works – just as well if the incompatibilist is allowed to keep their assumption that the agent in the deterministic setting most definitely is *not* responsible. This is just as well, for it seems to me that it *is* dialectically kosher for the incompatibilist to keep that assumption as the onus is on those who reject PAP to provide the arguments. There is, in other words, a standing presumption in favour of incompatibilism.

then one would expect that both the presence of determinism and the presence of the counterfactual intervener would be equally pertinent to an explanation of the agent's non-responsibility. In much the same way, if two assassins inflict fatal shot wounds on the prime minister at exactly the same time, it would be odd to explain the prime minister's death by citing the course of just one bullet, and ignoring the other.

Some have responded to this kind of case by pointing out that it is not clear that in the actual sequence it is the counterfactual intervener's presence that is doing any work ruling out the agent's alternative possibilities. It is the presence of determinism in the actual sequence that is extinguishing the alternatives, not the counterfactual intervener – hence there is a very natural explanation of their irrelevance to the agent's non-responsibility. To be convincing the case needs to be one in which the agent's alternative possibilities are ruled out solely by the counterfactual intervener (see Goetz 2005 *forthcoming;* Haji and McKenna 2004, p. 309; Ekstrom 2002, pp. 311-312). In other words, committed Frankfurtians must continue to explore variations such as those presented above by Mele and Robb, Stump and Pereboom.

This response to Fischer's case is understandable. But although I defend PAP, I do not think Fischer's argument can be dismissed quite so easily. Imagine that, per impossible, it *is* possible to construct a Frankfurt-style case in which the agent is indeterministically caused to decide but also lacks alternative possibilities due to the presence of a counterfactual intervener. We would have to agree that such cases provide a good piece of evidence for the falsity of PAP. Yet this evidence would stem from the apparent irrelevance of the counterfactual intervener's presence to the question of the agent's moral responsibility. This is exactly what Fischer's case draws our attention to as well. I grant that Fischer's case might not be as compelling as a case in which determinism is not assumed (if such a case could be devised). But nor can it be completely dismissed.[41]

However, I believe that Fischer's Frankfurt-style case, if it works, actually only puts pressure on a particular version of PAP. To see this we have to digress slightly.

[41] I make this point in my forthcoming c.

3.8 PAP and Ultimacy

In the previous chapter I pointed out that PAP is taken to encapsulate at least two demands. One is for genuine alternative possibilities in order to secure independence from the past and thereby be the ultimate 'source' of what we do, rather than just being a link in a chain. The second demand is for path-picking control, so that we do not just arbitrarily make a difference. The second concern is the one that I have been focussing on so far, and it has been the main focus of debates over free will.

However, just because the 'control' demand has been the most focussed on, does not mean that it is the most important or central. The demand that one not just be a link in a chain is also, on reflection, a powerful concern that could easily be the more fundamental. According to Sorabji, Aristotle asserts the UR condition when he says that "the concept of an action being up to us is connected... with the concept of our being, or having within us, the "origin" (*arche*) of the action" (1980, p. 234). Klein has defined the condition (which she calls the "Ultimacy-condition") as "the sense that nothing for which they are not responsible should be the source [or cause] of their decisions or choices" (1990, p. 51).[42] As Kane says, the UR condition "puts the emphasis for being up to us not on the power to do otherwise, but on the source or explanation of the action that is actually performed; that source must be "in us"'" (1996, p. 34). If determinism is true the UR condition is not met, for we are never the originators of our actions in a determined world, but are simply one link in a deterministic chain extending into the past.

If there is indeterminacy in our decision making processes – if, at the moment of choice it was genuinely possible for us to have chosen otherwise keeping all else constant – then we terminate the explanation of why we chose as we did rather than differently. Something, in other words, is ultimately down to us, and cannot be explained by citing factors external to ourselves. I am going to be saying much more about this kind of concern in coming chapters. The point to be recognised at the moment is that traditional Frankfurt-style cases, if they did work, would show that the demand for alternative possibilities could be separated from the demand for 'ultimacy' or 'sourcehood'. For in

[42] Klein does not think that we can ever be the true sources of our actions in the way required by the utlimacy condition (1990).

a traditional Frankfurt-style case the agent's decision making process is internally indeterministic such that it would, in the absence of an intervener, provide the agent with genuine alternative possibilities in which the agent decides otherwise. If the intervener's presence rules out such alternative possibilities without antecedently determining the agent's will, then the agent remains the ultimate source of their decision in the actual sequence, despite the lack of alternative possibilities. PAP and the Ultimacy condition would come apart.[43]

The failure of the traditional Frankfurt-style case means that there is *no* case for saying that PAP and Ultimacy come apart. Having genuine alternative possibilities remains intimately connected with being the ultimate source of our actions, and as such ultimacy remains a plausible part of the story behind the appeal of PAP. However, being the ultimate source of one's action *does not* require that one have the additional path-picking control associated with PAP. To stress – the requirement for path-picking control is a separate demand. Alternative possibilities are best seen as providing the space within which such control can be exercised, whereas alternative possibilities, just in themselves, can provide Ultimacy.

This means that we can distinguish two versions of PAP, one associated with the more minimal requirement for ultimacy, and one associated with the more demanding requirement for path-picking control.

PAP1: An agent is only morally responsible for making a particular decision if they had alternative possibilities in which they decided otherwise.

---

[43] It is important to recognise that for precisely this reason Frankfurt-style cases, if they can be got to work, do not directly threaten incompatibilist positions that stress ultimacy rather than control. There are a number of such incompatibilists, many of whom positively defend Frankfurt-style cases, including: Hunt 2000; Pereboom 1995 & 2000; Stump 1996; Zagzebski 2000. Some of these, in particular Pereboom, also thinks that in this way an incompatibilist who stresses control can also survive refutation by Frankfurt-style case, though as I indicated in an earlier footnote, it is mysterious to me how this could be. But I will not argue this point here. The more general point I would make is that I think that Frankfurt-style cases – if they worked – would put some kind of pressure on these positions, rather than supporting them (so I think the above would all be better advised to stop defending Frankfurt-style cases). For so long as ultimacy and PAP are strongly connected then provide mutual support for one another. Whereas if they are disconnected then it is far easier to call into question whether ultimacy is really all that valuable after all – something Fischer likes to do (1994, 2003b).

PAP2: An agent is only morally responsible for making a particular decision if they had alternative possibilities in which they decided otherwise, and path-picking control over which possibility was actualised.

Now we can return to Fischer's Frankfurt-style case. Consider that the traditional Frankfurt-style case – if it could be got to work – would have refuted both PAP1 and PAP2. This is because the traditional Frankfurt-style case presents us with an agent who is intuitively morally responsible yet lacked alternative possibilities – something both PAP1 and PAP2 say is impossible. But I claim that what is interesting about Fischer's case, is that it only threatens PAP2, if it threatens anything at all.

Let us call an incompatibilist who puts the stress on ultimacy a 'source' incompatibilist.[44] PAP1 is associated with source incompatibilism. Let us call an incompatibilist who puts the stress on path-picking control, a 'control' incompatibilist.[45] PAP2 is associated with 'control' incompatibilism. In Fischer's case what both the 'source' and the 'control' incompatibilist need to be able to do is explain the intuitive irrelevance of the counterfactual intervener. If they can do this, then the case is no threat to their respective positions.

The 'control' incompatibilist cannot do this because alternative possibilities are needed by the control incompatibilist in order to clear room to facilitate the exercise of their brand of control. Because both the counterfactual intervener *and* the determinism in the actual sequence are ruling out alternative possibilities, then both are obstacles to the exercise of this kind of control. One is no *more* of an obstacle than the other. Thus the control-incompatibilist cannot just focus on the determinism in their explanation of the agent's non-responsibility.

But matters are different for the 'source' incompatibilist. This is because for the source incompatibilist genuine alternative possibilities are the upshot of one being the ultimate source of one's action. Alternative possibilities are not clearing room for the exercise of some ability. They are not making space. Alternative possibilities are just what an agent

[44] With some qualifications to be made later, contemporary source incompatibilists would include: Wiggins 2003; Nozick 1981; Kane 1996; Ekstrom 2000..

[45] Prominent 'control' incompatibilists would include: van Inwagen 1983 & 2002; Clarke 1993 & 1996; O'Connor 1995; Rowe 2000; Taylor 1992.

will have if they are the ultimate source of their action. This means that the source incompatibilist *can* give a very natural explanation of why it is that the counterfactual intervener's presence seems irrelevant to the agent's non-responsibility. For it is in virtue of the antecedent determination of the agent's will that the agent is not the ultimate source of their decision. So the determinism is the culprit, and not the counterfactual intervener. The determinism is what is extinguishing the agent's sourcehood in the actual sequence. It is precisely because alternative possibilities would just be the incidental upshot of one having sourcehood that the intervener's presence can be deemed irrelevant.

There is another way to make the point. If one were to remove the determinism and make the agent's deliberative process indeterminate, then the agent would be the ultimate source of their action and would have genuine alternative possibilities (it would take a viable traditional Frankfurt-style case to show the second part of this claim to be false). But this is not so for the control incompatibilist. For simply removing the determinism would not, in itself, yield path-picking control (it is true that some contemporary 'event causal' incompatibilists do seem to think that it might, but I take issue with them on that score in Chapter 5). Path-picking control is something needed *in addition* to having genuine alternative possibilities. This brings out that for the control incompatibilist alternative possibilities have to be seen as facilitating the exercise of path-picking control, whereas with the source incompatibilist alternative possibilities are the upshot of the sourcehood condition being satisfied.

If I am right above this is significant. It means that not all PAPs are the same. It means that Fischer's case puts pressure on the ability or control aspect of PAP, but not its 'source' aspect. In other words, Fischer's case calls into question why PAP is a requirement of moral responsibility, rather than whether it is. Most importantly of all, given the view I will be developing in coming chapters, Fischer's case suggests that the reason PAP is a requirement of the kind of free will associated with moral responsibility, is to do with being an ultimate source of what one does, rather than having a special kind of path-picking control over what one does.

3.9 Objections to the Above

One might object that even if I am right above, this does not really matter given that it is precisely the control version of PAP that the compatibilist is interested in dismissing. After all, if Fischer's case calls into question whether we actually need such control, then compatibilist control seems to be the only kind needed for responsibility. As such ultimate luck cannot be deemed the problem it is thought to be (and as we shall see later (in Chapter 8) Fischer does directly challenge whether ultimate luck is such a problem). So, what I have argued above will not trouble the compatibilist.

There is some merit in this response. If the compatibilist's concern is just to show that compatibilist control is the only kind actually needed for responsibility then I believe that Fischer's case does provide some evidence for this. Hardly compelling evidence, but evidence none-the-less. But there is a difference between being a compatibilist about control, and being a compatibilist about free will and responsibility. It is perfectly consistent to be the former but not the latter, for free will and responsibility are not plausibly only about control (for instance, sanity is in general a control independent requirement for responsibility and is neutral between compatibilists and incompatibilists). The concern about ultimate sourcehood is a real concern and not one the compatibilist can just dismiss.[46] It is not, after all, a concern that has just been invented to be an obstacle to compatibilism. It is not, in other words, an ad hoc reason to be an incompatibilist. A compatibilist is not allowed to just *dismiss* this kind of concern about determinism in the absence of any argument. I would agree, that if traditional

---

[46] Fischer himself often does dismiss the concern about sourcehood or at least seems to. He says things like this:

> Someone might say that in order for an agent to be morally responsible for an action, the agent must be creative in the sense of being a "self-initiator" or "self-originator" of the action. And the claim would be that these ideas require the absence of causal determination. Now I can see why someone might insist that responsibility requires this sort of incompatibilistic creativity, *if* one is committed to the idea that moral responsibility requires alternative possibilities, but I do not see any reason to insist on precisely *this* sort of creativity, *apart* from such a prior commitment. (1994, p. 150)

In other words, he fails to see the value in sourcehood once it is divorced from the concern about alternative possibilities. I have argued here that it cannot be so divorced. But anyway, I find it hard to understand exactly what Fischer means when he says this. Does he mean that sourcehood could only be valuable if it was associated with having genuine alternative possibilities? If *that* is what he means, then I am inclined to agree to this extent – the intuitive requirement for alternative possibilities can be given a rationalisation in terms of sourcehood. But I do not understand why Fischer can grant that there is a reason to insist upon "self-origination" when it is associated with alternative possibilities, yet no reason at all, when it is shown (as he thinks it can be) to be divorced from alternative possibilities.

Frankfurt-style cases could be got to work, then the sourcehood requirement and PAP would be demonstrated to come apart. This, I would agree, would *weaken* the case for source-incompatibilism insofar as the requirement for sourcehood could no longer be associated with (and so be part of the story behind) the intuitively appealing PAP. But if what I have argued here is correct, then PAP and sourcehood cannot be divorced. Furthermore, to the extent that Fischer Frankfurt-style cases work, then PAP and sourcehood are shown to be more closely linked than before. For I've argued that Fischer cases actually suggest that sourcehood is the real story behind the requirement for alternative possibilities.

3.10 Conclusion

By challenging PAP Frankfurt-style cases attempt to demonstrate that whatever value we might attach to having an open future, and whatever such an open future facilitates, it is all irrelevant when it comes to the question of moral responsibly. An ability to decide otherwise is not needed, and it is enough if the agent has exercised just some form of compatibilist control over their decision in the actual sequence. However, in this chapter I have argued that PAP cannot be so easily dismissed. The arguments of Widerker and Kane show the original Frankfurt-style case does not present us with a situation in which an agent lacks genuine alternative possibilities, and hence our intuitions about such cases provide no evidence for the falsity of PAP.

Subsequent attempts to repair the damage done by Widerker and Kane's criticisms, all fail. The only version that does any work against PAP is Fischer's determinism-assuming variation. But Fischer's version is different from the others and requires quite careful treatment. I agreed that whilst Fischer's version is a long way short of decisive, it does provide some evidence that an ability to do otherwise might not be required for responsibility. However, PAP can be said to encapsulate at least two concerns about determinism, only one of which is a concern about control. I argued that PAP, insofar as it encapsulates a concern about being the ultimate source of one's action, is *not* threatened in any way by Fischer's case. At most Fischer's case serves only to aide our understanding of PAP, rather than serving to provide us with a reason to reject it.

If a Frankfurt-style case can call into question any plausible interpretation of PAP, then this is good news for the compatibilist. If Fischer's case calls into question whether the ability to decide otherwise – so path-picking control – is really needed for responsibility, then Fischer's case provides some evidence that luck at the ultimate level might not be such a problem. That is all well and good for the compatibilist, because if determinism is true, then everything we do is a matter of ultimate luck. Fischer's Frankfurt-style case can therefore be seen as going some small way to alleviating concerns about luck and the adequacy of compatibilist control.

However, compatibilist positions are not only subject to concerns about the adequacy of compatibilist control. As we shall see in the next chapter, cases involving covert manipulation of an agent's will provide a powerful case against compatibilism.

Chapter 4

Compatibilism and CNC Control

In the previous chapter we considered whether Frankfurt-style cases provide any evidence that PAP is false. I argued that they do not, and rebutted three of the latest 'repaired' Frankfurt-style cases. I did allow that Fischer's variation provided some evidence that the demand for path-picking control might be misguided. But if that is right, then even though Frankfurt-style cases cannot show all that some have claimed for them, they do improve the prospects for compatibilism about control.

However, compatibilist views about free will have *two* Achilles' heels. There's the luck objection, which attacks the adequacy of compatibilist control but, there is also the objection one gets from cases involving "covert non-constraining" control or CNC control.[47] The term was coined by Kane, who explains that CNC "controllers do not get their way by constraining or coercing others against their wills, but rather by manipulation the wills of others so that the others (willingly) do what the controllers desire" (Kane 1996, pp. 64-65).

The argument goes that in cases in which an agent is subject to CNC control their responsibility is clearly subverted. Yet there is no relevant difference between such cases and those involving the antecedent determination of the agent's will. In other words, the reason CNC control undermines responsibility is because it involves the antecedent determination of the agent's will. It is this line of attack that I will pursue in this chapter. My argument develops in the following manner.

Many compatibilists accept that CNC control subverts responsibility and pick up the gauntlet of trying to explain why without implicating determinism.[48] Because CNC controllers do not jeopardise the agent's control, the compatibilist has to acknowledge

[47] For discussion of CNC control see: Waller 1990; Double 1991; Mele 1995; Pereboom 1995; Fischer & Ravizza 1998, chs. 7 & 8; Haji 1998 pp. 21-25 & 2002 ch. 8.

[48] Fischer and Ravizza (1998) and Haji (1998 & 2002) have been prominent amongst those compatibilists who accept that CNC control presents a problem and try to deal with it. Harry Frankfurt has also struggled with the issue, though he now bites the bullet and just insists that an agent *is* morally responsible even if they are subject to CNC control. The compatibilist Gary Watson honestly accepts that CNC control presents a major problem for compatibilists without offering a solution to it.

that it is the violation of some kind of 'ownership' or 'deep attributability' condition which undermines the agent's responsibility. When an agent is subject to CNC control their actions, though still compatibilist controlled, are no longer really *theirs*. I then consider various compatibilist attempts to articulate what is needed for ownership. I find that in each case either the conditions outlined fail to preclude being CNC controller, or amount to simply stipulating that CNC control subverts ownership. Finally, I argue that Frankfurt-style cases can actually be used to help underscore the lack of relevant distinction between the responsibility subverting qualities of CNC control, and causal determinism. In short cases involving CNC control motivate an ownership condition for moral responsibility, and compatibilist articulations of this condition fail.

4.1 Coming To Be Who We Are

How we come to be the kinds of people that we are, is a process over which we have not, ultimately, exercised control. We do not come into the world already formed but rather "from the very beginning that process [of self-creation] is given its own distinctive slant by the influences of heredity and environment." (Feinberg 1986 p. 34). As Feinberg goes on to say:

> At a time so early that the questions of how to socialize and educate the child have not even arisen yet, the twig will be bent in a certain definite direction... From the very beginning, then, the child must – inevitably *will* – have some input in his own shaping, the extent of which will grow continuously even as the child's character itself does. After that, the child can contribute towards the making of his own self and circumstances in ever increasing degree. These contributions are significant even though the child is in large part... the product of external influences over which he has no control, and his original motivational structure is something he just finds himself with, not something he consciously creates. Always the self that contributes to the making of the newer self is the product both of outside influences *and* an earlier self that was not quite as fully formed. (1986, pp. 34-35; see also Arneson 1994, pp. 59-60)

The original self is not consciously self-created, and how it develops will be a function both of it, and the influences to which it is subjected. Yet Feinberg does not think this

undermines free will. For the background character traits and evaluative schemes that grow through this process of maturation can still be part of an 'authentic' self, if "the habit of critical self-revision was implanted in us early by parents, educators, or peers, and strengthened by our own constant exercise of it" (Feinberg 1986, p. 35). In other words, although we have not controlled our acquisition of the disposition to engage in critical self-revision, this fact does not in any way subvert our free will and responsibility, but rather can be seen as free will enabling.

These are important points which both incompatibilists and compatibilists have to recognise. Even the most hard-line incompatibilist has to accept that we do not just appear on the scene ready made, but develop over time, and furthermore that we need certain dispositions installed if we are to become responsible beings. The habit of self assessment and some kind of normative competence both seem to be requirements here. But the point the incompatibilist would make, however, is that it is not just the having of the disposition to self-revise that is important, it is rather what this disposition facilitates. For the habit of critical self-assessment will likely involve seeing that one has alternative possibilities – that to some extent it is up to oneself now how one develops. In other word, such a habit can be seen as in effect generating the alternative possibilities that we need if we are to be the originators of our actions, and if we are to exercise path-picking control over them.

The problem the compatibilist faces, however, and which is going to be the focus of this chapter, is as follows. Whilst we can distinguish between free will enabling, and free will disabling dispositions, the problem for the compatibilist is that a sophisticated covert controller could programme an agent to perform certain future acts, consistent with the agent critically reflecting upon what they are doing and possessing sufficient normative competence. In such cases the agent is subject to what has become known as covert non-constraining control, or CNC control. CNC control is to be contrasted with constraining control. An agent is subject to constraining control when a controller forces them to do something against their will, or places some obstacle in the way of their doing what they want. In short "[c]onstraining... controllers get their way by creating constraints or impediments that thwart the wills of those they control, preventing other agents from doing what they want to do" (Kane 1996, p. 64). Whereas in cases of non-constraining control, "the controllers do not get their way by constraining or coercing

others against their wills, but rather by manipulating the wills of others s that the others (willingly) do what the controllers desire" (Kane 1996, p. 64).

> [CNC controlled] agents consequently do not feel frustrated or thwarted. They act in accordance with their own wants, desires or intentions. Yet they are controlled nevertheless by others who have manipulated their circumstances so that they want, desire, or intend only what the controllers have planned. In the most interesting cases... the controlled agents are unaware of being manipulated or perhaps even unaware of the existence of their controllers. (Kane 1996, pp. 64-65)

We can further distinguish between two types of CNC control: Global and Local. Global CNC control involves the wholesale implantation of the agent's actional elements, so the agent's psychological basis for decision making and their background evaluative scheme. These might be implanted in embryo form, so to speak, the CNC controller foreseeing that the agent will, over time, develop a certain background evaluative scheme – one which will ensure the agent acts in the way the CNC controller wants them to. Global CNC control is of a kind that a genetic manipulator of the future might exercise, or it may be the kind of control that a god exercises over us.[49]

Local CNC control does not involve the wholesale construction or reconstruction of the agent's control mechanisms, rather it involves inducing certain desires and beliefs in the agent who is otherwise left intact. Local manipulation takes place when a controller induces in the agent a particular desire which they know will elicit a particular reaction from the agent's 'in place' mechanism. Thus, whereas Global manipulation involves the implantation or engineering of the agent's control mechanism, local manipulation involves presenting the already formed mechanism with circumstances (and here I include induced pro-attitudes as part of an agent's circumstances) in which it will react in the way that the local controller intends. It is possible that an agent can be globally controlled without ever being locally covertly controlled and vice versa. Global control involves pre-arranging matters so that an agent's future is fixed and will be what the

[49] This kind of global CNC control is also illustrated by *Brave New World* type behaviour engineering. Gary Watson uses the term "Brave New World scenarios" to mean global CNC cases. But as Kane points out "CNC control itself is a more general notion than global scenarios suggest, however, since it can take

global controller wants it to be. The global controller simply winds up their carefully engineered clockwork toy and releases it. They can sit back; they do not have to actively intervene after their initial act of wholesale manipulation. The local controller on the other hand intervenes, but not by bypassing the agent's reason responsive apparatus (again, that would be constraining local control), but rather by presenting it with incentives that it knows will elicit the desired reaction.

There is some ambiguity over when exactly local CNC control shades over into Global. For instance, the local manipulation of the agent's evaluative scheme (so the basis upon which they assess the reasons provided by their desires etc) arguably constitutes Global manipulation. We might say that an agent just *is* their evaluative scheme, or at least that it is in some way essential to them, and to start playing around with that, is to start changing the agent's essential identity.

The important point about these types of covert control, whether global or local, is that in all events, what the agent does results from the operation of reason-responsive mechanisms, and therefore in terms of compatibilist-control, the agent can be said to have controlled their decision. That is the problem.

For as I mentioned in Chapter 2, there are some things which just should undermine moral responsibility, and being subject to CNC control is plausibly one of them. If everything we do is a function of what some CNC controller wants us to do, then intuitively we are *not* morally responsible. That, at least, is the intuition of many. If the compatibilist wants to respect such intuitions – and many do – then they have to explain exactly *why* one is not morally responsible in these cases. More specifically, they need to explain this in a way that will not show determinism to be the culprit.

4.2 Pereboom's Four Case Argument

To bring out the problem consider the following 'four case' argument developed by Derk Pereboom:

---

place to various degrees in everyday life as well as globally in utopian scenarios like Huxley's" (1996, p. 65 fn. 9).

> Case 1: Mr. Green is like an ordinary human being, except that he was created by neuroscientists, who can manipulate him directly through the use of radio-like technology...The neuroscientists manipulate him by, among other things, pushing a series of buttons just before he begins to reason about his situation, thereby causing his reasoning process to be rationally egoistic. His reasoning process is reasons-responsive, because it would have resulted in different choices in some situations in which the egoistic reasons were otherwise. Mr Green does not think and act contrary to character, since the neuroscientists typically manipulate him to be rationally egoistic.
>
> Case 2: Mr. Green is like an ordinary human being, except that he was created by neuroscientists, who, although they cannot control him directly, have programmed him to be a rational egoist, so that, in any circumstances like those in which he now finds himself, he is causally determined to undertake the reasons-responsive process and to possess the set of first and second-order desires that results in his killing Ms. Peacock.
>
> Case 3: Mr. Green is an ordinary human being, except that he was determined by the rigorous training practices of his home and community to be a rational egoist. His training took place at too early an agent for him to have had the ability to prevent or alter the practices that determined his character. Mr. Green is thereby caused to undertake the reasons-responsive process and to possess the organisation of first and second –order desires that result in his killing Ms. Peacock.
>
> Case 4: Physicalist determinism is true. Mr. Green is a rationally egoistic but (otherwise) ordinary human being, raised in normal circumstances. Mr. Green's killing of Ms. Peacock comes about as a result of his undertaking the reasons-responsive process of deliberation, and he has the specified organization of first and second-order desires. (Pereboom 1995, pp. 23-25)

Pereboom thinks most of us will have the intuition that in cases 1,2, and 3 Mr Green's responsibility for his act of killing has been subverted, and that case 4 is relevantly identical to cases 1,2 and 3, and thus we should conclude that Mr Green is not

responsible in case 4 either. The reason why Mr Green is not responsible in the first three cases is because what he does "results from a deterministic causal process that traces back to factors beyond his control", which is also what one has in case 4 (Pereboom 1995, p. 25).

The moral of CNC cases is incompatibilist. It seems that the incompatibilist can, it would seem, deal CNC control cases straightforwardly. The control incompatibilist can point out that if an agent has path-picking control, then even if all of their actional elements have been implanted, the agent will still have ultimate control over what they do. The CNC controller will not, through implanting the agent's actional elements, be able to ensure that the agent *freely* decides one way rather than another on a particular occasion, for by hypothesis this will be a function of how the agent's path-picking control is exercised – a matter that cannot be antecedently determined. Possession of path-picking control therefore effectively thwarts CNC control. Similarly, the source incompatibilist can claim that meeting the source requirements rules out a CNC controller being able to determine which way an agent will decide. Either way, for one reason or another the CNC controllers power over the agent is reduced if the incompatibilist control/sourcehood requirements are met. Whether the incompatibilist can actually deal with CNC control cases quite as easily as I've just implied is a matter for discussion – some think not (Haji 2002; Mele 1995). It is a discussion that will occupy part of Chapter 7 where we will see how some incompatibilist positions fare better than others, and that all require certain adjustments. For the time being, however, we can just note that the incompatibilist looks far better equipped to deal with cases involving CNC manipulation.

4.3 CNC Control and Ownership

With a few notable exceptions, most contemporary compatibilists agree that cases involving CNC control do subvert responsibility, and so therefore pick up the gauntlet of trying to explain why, without implicating determinism. Because CNC controllers go through the agent's will so to speak, they do not subvert the agent's compatibilist control. Whether one does what one does as a result of the prior manipulations of a CNC controller, or whether one doe what one does as a result of natural forces, it makes no difference to the degree of compatibilist control one exercises (anymore than it

affects the degree of control a television remote control can be said to exhibit whether it was assembled in a factory or assembled by natural forces). So instead of focussing on control, the compatibilist will have to say that what CNC control attacks is the agent's ownership over what they do. I borrow the term ownership from Fischer, and it is my preferred term, but others are referring to essentially the same condition when they talk about attributability or autonomy. The uniting idea is that in some sense the agent's act does not truly belong to them in CNC cases.

We have seen in previous chapters that this is a condition that incompatibilists prize too – it is one part of the story behind PAP. But just as there is disagreement over whether compatibilist control is adequate for moral responsibility, so too there is room for disagreement over whether determinism provides circumstances in which we can have a suitable form of ownership over what we do. In what follows I will consider various compatibilist accounts of ownership, and will find them all wanting.[50]

4.4 Hierarchical Accounts of Ownership

A boundary condition on an acceptable account of 'ownership' must be that it is inconsistent with being subject to CNC control. For a compatibilist has to be able to say that the reason CNC control subverts responsibility is that it violates the ownership condition.

In this section I will argue that acknowledging this point rules out what are known as

---

[50] It might be suggested that what really concerns us about CNC control is the thought that we are subject to the manipulations of those who do not have our best interests at heart. In other words, it is not *that* we are CNC controlled that is the problem, but rather that CNC controllers are unlikely to have our best interests at heart (Waller 1988, p. 165).
However, this is unconvincing for two reasons. Firstly, it is not obvious how this distinguishes CNC control from determinism. After all, in what sense does nature have our best interests at heart? Why, in other words, shouldn't our concerns about determinism be exactly akin to our concerns about CNC control? Secondly, we are likely to feel CNC subverts responsibility even when it has been exercised in our interests. After all, the reflection that a benevolent God is controlling all that we do is still, to most of us, objectionable. Kane uses a more prosaic but telling example:

> The problem of benign CNC control is also evident in the case of children who may know that their parents are well-intentioned toward them, yet they resent parental autonomy anyway. As they reach their maturity the children want something over and above the removal of fear that their parents will choose badly for them. They want an autonomy and dignity that they associate with the power to run their own lives, to make and take responsibility for their own choices even at the expense of making mistakes and being worse off than if they had let their parents run their lives. (1996, p. 69)

hierarchical accounts of ownership of the kind associated with Frankfurt, Watson and Taylor (Frankfurt 1971; Watson 1975; Taylor 1976). According to a hierarchical account what matters is that the agent has first order desires that are endorsed by their second order desires; Jones wants to do X, and he wants to want to do X. In this case we have the appropriate mesh between first and second order desires, and as such Jones can be said to have ownership over his doing X. He has secured "the conformity of his will to his second-order volitions" (Frankfurt 1971, p. 16). This is to be contrasted with cases in which an agent has a first order desire that is not endorsed by a second order desire. Jones might want a cigarette, but not want to want a cigarette. In these kinds of case Jones's will is not free, for Jones does not take ownership of what he is doing. From Jones's perspective his subsequent smoking is going to strike him as being as much something that is happening to him, as something that he is doing. This is how things appear when first and second order desires are not aligned. You do things that you want to do, but you do not want to want the things you want. On this hierarchical view ownership is about internal harmony between one's first and second order desires (and however many other orders of desires one has). We can see how this kind of view is often effective at delivering the correct verdict about victims of brainwashing, addicts, obsessions, phobias and other neuroses. For though such people might do what they want, they will typically not second order endorse their first order wants. The agoraphobic does not want to go out, but they do want to want to go outside.

We should, I think, agree that this kind of analysis is insightful and something that a full account of ownership needs to include. A plausible account of responsibility must involve "the ability to step back and ask ourselves whether or not we should act on our various motivations and desires; to attain some critical distance from them and choose which to endorse, rather than acting on them unreflectively and accepting them uncritically" (Bok 2003, p. 162). Nevertheless the ability to act on our second order desires cannot constitute a full account of the kind of ownership required for responsibility[51] For the problem is simply that the global CNC controller can see to it that the agent second-order endorses all that they do. For whether the appropriate mesh between first and second order desires is a matter of the current time slice properties of the agent's mechanism, and has nothing to do with how the mesh came to obtain –

nothing, in other words, to do with its history. So, if second-order endorsement is the full story about ownership it will fail to preclude global covert control.

> If we came to [secure conformity between second and first order desires] not by our own efforts but by brainwashing or being manipulated by a cult leader, or conditioned by behavioural engineers, or given a powerful mind-altering drug, we might be satisfied with the will we had, but would lack autonomy. (Kane 1996, p. 64)

Harry Frankfurt has himself struggled with the issue of ownership. He has suggested that in a CNC case where a CNC controller continuously manipulates an agent – so a case of local CNC control - then the agent's responsibility is subverted for the following reason:

> [T]he subject is not a person at all. His history is utterly episodic and without inherent connectedness. Whatever identifiable themes it may reveal are not internally rooted; they cannot be understood as constitution or belonging to the subject's own nature. Rather, they are provided gratuitously by an agency external to the subject. (1988, p. 53)

However, in a case of global CNC manipulation, where the agent is provided with a stable character Frankfurt holds that the agent is morally responsible for what he subsequently does, provided that he identifies "himself with some of his own second-order desires, so that they re not merely desires that he happens to have or to find within himself, but desires that he adopts or puts himself behind" (Frankfurt 1988, p. 53).

In the first case, Frankfurt seems to be supposing that local CNC manipulation will necessarily result in a discordant agent history without connectedness. This is false. Nothing prevents a local CNC controller manipulating an agent in a way that preserves or continues or even brings the "continuity and intelligibility essential to being a person" (1988, p. 53). Furthermore, by Frankfurt's own lights his judgement regarding the agent's responsibility in this kind of case is not warranted. As Haji puts it, "[t]he fact

---

[51] Those who accept the power of reflection associated which hierarchical accounts make central include: Neely 1974; Benson 1987; Gert and Duggan 1979; Davis 1979; Young 1979; Zimmerman 1981; Taylor

that identification is not "internally rooted" is *irrelevant,* given a history insensitive view of the sort of agency presupposed by responsibility, to the status of an agent as a person. Frankfurt's remarks that the victim's having "no character or dispositions of *his own*" (emphasis added) undercut his own view that one "makes" one's desires one's *own* by identifying with them, on the supposition that identification is history insensitive" (2002, p. 134).

Note also that an agent's history might be "episodic and without inherent connectedness" naturally and not as a result of manipulation. It is not at all obvious that such an agent would fail to be responsible for the decisions they actually make where control conditions are satisfied (as, by hypothesis, they are). It is true, that there is room for disagreement here. Historically Hume held that an agent "is not answerable for [immoral acts if] they proceeded from nothing in him, that is durable and constant" (1955, p. 98). There must come a point where a lack of 'inherent connectedness' would lead to a lack of responsibility. But the fact remains that CNC control is not essentially related to having an episodic history that lacks connectedness, and we do not always consider a lack of connectedness to subvert responsibility (if an agent acts *out of character* for instance, we still might consider them blameworthy, provided we think that they controlled what they did sufficiently). This makes it far more natural to conclude that the agent's responsibility is subverted in this kind of case because they are subject to CNC control, and not for reasons to do with the stability of the agent's character.[52]

In the second case, Frankfurt's point is that it does not matter how one came to be the way that one is. If one identifies with the way that one is, one is responsible – for one has *taken* responsibility. I think this sounds half-way plausible only if manipulation is assumed to have taken place at the first order level and the agent comes to second-order endorse what is going on at that level (even here I think our intuitions are not clear).[53]

1976; Dworkin 1970 & 1988; Velleman 1989; Zimmerman 1989; Kapitan 1989; Kane 1996 p. 66.

[52] I say that the agent would be responsible where their discordant character had been acquired as a result of a natural process, but needless to say I maintain that in deterministic settings this would not be so. That prejudges the outcome of the arguments I present here, so I allow for the sake of argument that determinism does not undermine responsibility until we have reason to think otherwise.

[53] Note also that even if this were somehow precluded, the resulting view is still quite unattractive. For imagine that the manipulated agent realises that their volitions have been externally implanted, but nevertheless "identifies" with them, then on this view the agent is responsible. It is characteristic of the hierarchical model of ownership that "acting freely... is a matter of resigning and adapting oneself to necessity" (Zimmerman 2000, p. 25).

But nothing prevents the CNC manipulator from having engineered at the second order level too (Haji 2002, p. 135). In other words, the manipulator has designed the agent in such a way as to ensure that they will second-order endorse the way that they are and *take* responsibility.

> [Y]ou could imagine any degree of reflectiveness and self-awareness you wish. You may imagine persons so sophisticated that they are thinking about the appropriateness of their fourth – or fifth – or twentieth –order desires. And yet they still might be CNC controlled by more sophisticated beings who have manipulated their highest order desires. (At the very highest level of sophistication one might imagine God doing the controlling, since the problems posed by CNC control have their theological counterparts in problems of divine predestination or foreordination.) (Kane 1996, p. 66)

Thus it does not matter how many levels one adds, or how complicated one makes the set-up, one can always introduce a more sophisticated CNC controller to deal with the extra complication. There is no way of 'building' one's way out of this problem by making person's ever more reflective.[54]

It is surprising how many philosophers seem to think that the subjective 'taking' of responsibility is relevant to the question of whether one actually *is* responsible. We see that Frankfurt thinks that this is important, but so do Fischer and Ravizza, Haji, and G. Strawson. Yet it seems to me entirely irrelevant for two reasons. First for the reason given above: our mental act of 'taking responsibility' is as susceptible to CNC manipulation as anything else. Second, because we can fail to take responsibility, yet still be morally responsible. This can be illustrated in the following way. Most of us, if we came to believe that we were subject to CNC manipulation, would *not* take responsibility. Imagine that, taking this on board, some of us decide that because we now have an excuse for all that we do, there is no reason not to behave appallingly. We duly do so. We then discover new evidence which suggests that we are not subject to CNC control after all. I have the intuition here that we are morally responsible for those

[54] It is not even clear how reflectiveness even begins to help (which is certainly not to say that reflectiveness is not required for responsibility). As Kane notes, "sometimes reflective persons are susceptible to manipulation precisely because they are reflective, so long as others know enough to predict the lines their reflection will take" (1996, pp. 66-67).

acts that we committed whilst believing ourselves to be subject to CNC control. In other words, I have the intuition that the belief in CNC control does not provide one with an excuse. If this is right it brings out the objective dimension to free will and moral responsibility. Whether one acts freely is not a matter determined solely by focussing on the agent's subjective states. For there is no difference, in terms of subjective states, between someone who has the justified but false belief that they are subject to CNC control and someone who has the justified true belief that they are subject to such control. Yet there *is* a difference in their free will and moral responsibility.[55]

4.5 Historical Accounts

Given what has been said so far, it looks as if a compatibilist account of ownership is going to have to be historical. It is going to have to make it matter how the agent came to be the way they are, and not just focus on how the agent is now. By making certain historical facts important, this brings in an objective dimension. With this in mind we can now look at the compatibilist account of ownership offered by Fischer and Ravizza.

Fischer and Ravizza give a historical account of ownership according to which ownership is achieved as a result of going through a certain process in which the agent comes to "take responsibility" for their mechanism (1998, p. 200). As Fischer explains:

> [O]n the approach to compatibilism I favour, one looks carefully at the history of the behaviour in question. If there is unconsented-to covert manipulation of certain sorts, this can be the sort of historical factor that rules out moral responsibility. On my approach, one demands that the behaviour issue from the agent's own suitably reason-sensitive mechanism. That is, the agent must - in a specified sense – have "ownership" of the process that leads to the behaviour... These conditions are not met in the objectionable cases of [CNC control], and yet I would argue that they can be met in a context of mere causal determination.

---

[55] If I am right here, then this brings out something else too. Even if we believe that we are subject to CNC control, it remains rational to continue *as if* one is not, just in case one's belief is false. And the same holds true with respect to the truth of determinism. The incompatibilist has good reason to continue *as if* determinism is false, even in the face of good evidence for its truth, for it is always *possible* that one's belief is false – and it would afford one no excuse if this turned out to be the case. I will say more about this in the final chapter, in a section where I turn attention to the 'existence' question and the relative unlikelihood of incompatibilist free will. For the time being it is enough that we note how CNC cases draw attention to the fact that free will has an *objective* dimension.

(Fischer 2003b, p. 203) [56]

Fischer and Ravizza outline three necessary and sufficient conditions for taking responsibility or achieving ownership. The first two stipulate that taking responsibility involves first coming to see oneself "as an agent" (1998, p. 208; pp. 210-211; p. 238). That is to say that we see ourselves as being agents of change – as bringing about certain outcomes in the world. Once one sees oneself as an agent the way is clear for us to satisfy the second condition which involves seeing ourselves as "a fair target for the reactive attitudes as a result of how [we exercise] this agency in certain contexts" (Fischer and Ravizza 1998, p. 211). As has been pointed out earlier, these are the kinds of conditions that could be satisfied by way of covert manipulation. Nothing prevents a sophisticated enough controller engineering us such that we would go through this process. It is the third condition which is supposed to rule this out: "the cluster of beliefs specified by the first two conditions must be based, in an appropriate way, on the individual's evidence" (1998, p. 238). So, the agent must "have a certain kind of view of himself, in order to be morally responsible for his behaviour", he must "see himself as an agent who is an appropriate candidate for the reactive attitudes" (1998, pp. 220-3; p. 229).

But as Russell points out, "[a]t this critical juncture, however, the argument... seems to run out of line... the reader is asked to accept that there is an intuitive distinction between appropriate and inappropriate ways that an agent comes to see himself as an agent and a fair target of reactive attitudes" (2002a, p. 598). Fischer and Ravizza simply insist that "the relevant notion of appropriateness must remain unanalyzed" (1998, p. 236). Nevertheless they feel confident that we can agree that it is appropriate that what one does can be determined by natural forces, but not by the artificial forces encapsulated by covert controllers of various kinds.

---

[56] Fischer's claim that 'unconsented-to manipulation' poses the problem is misleading. Although the control in question is unconsented-to, it involves the implantation of the consent-enabling mechanism. To this extent the alternative – that one's mechanism be acquired by some natural process – is also not one to which the agent could consent. The only possible way in which such consent could be achieved is post the process in question. The covert controller could have so engineered things such that one would, retrospectively consent to the implantation process. Talk of consent is therefore unhelpful, for it implies that the problem stems from the agent's inability to control the process by which they come to acquire their reason-responsive mechanism, and it is not an option for a compatibilist to see the problem in this way. If they did, they would have to presuppose that control is prior to ownership, and this would only rule out covert control of the type in question if the agent is presupposed to have some kind of libertarian control.

As an attempt to circumvent the problem of covert control Fischer and Ravizza's account comes down to the claim that covert control undermines moral responsibility, but determination by nature does not. They just want us to see that one robs the agent of ownership whilst the other does not. Ultimately the conditions they outline do not explain why this is so and we just have an appeal to our intuitions on the crucial issue we want explained. This, I believe, is not going to be persuasive to anyone not already convinced of the truth of compatibilism. So, to the legitimate concern about the relevant difference between CNC control and causal determinism Fischer and Ravizza have nothing to say.

4.6 Dennett

I will now consider Dennett's approach.[57] Dennett asks us to distinguish between control by purposeful agents, and determination by nature. He then insists that it is this distinction which explains why in cases of covert control the agent is not morally responsible. When an agent is subject to covert control they are subject to someone else's project: the project of a purposeful agent. If they had instead been subject only to natural determination then, because nature has no purpose, they would not have been subject to control.[58]

In response, firstly I believe the distinction Dennett draws to be hollow. As Kane points out, when it comes to 'constraining' control, it makes no difference to my freedom whether I am constrained by natural forces, or by a purposeful, controlling, agent. It makes no difference to my freedom, for instance, whether the wind slammed the door shut locking me in the room, or whether Larry did. So it is odd, to say the least, that

[57] Dennett initially tries to do what I have just dismissed as doomed to failure above. Namely he argue that as long as the agent's control mechanism is of complexity sufficient for "self monitoring" then it will be able to detect and undo "the process of conditioning" (Dennett 1984, pp. 33-34). How though, when the controller will have controlled the form this self-monitoring will take? Dennett is here trying to 'build' his way out of trouble. I have already said what I want to about those who try to do this and so ignore this aspect of Dennett's approach here.

[58] I find it odd that Dennett should have any problem with CNC manipulation at all. For Dennett is a utilitarian about responsibility. In other words, he rejects that we ever truly *deserve* praise or blame, punishment or desert. Rather our blaming and praising practices are just about attempts to *manipulate* one another into conforming to standards which , in general, will maximise utility. In other words, Dennett thinks we *should* be manipulated, and that this is what our practices are all about (with the proviso that our practices might need revising in order that we should be manipulated more effectively) (see Dennett 1984 esp. pp. 139-144; and also Russell 2002b for elaboration of this point, pp. 246-247).

what is entirely irrelevant when it comes to constraining control should make such a significant difference when it comes to CNC control. This is a point that Kane has made nicely below:

> Your power or freedom to run or dance is no less impaired if you are paralysed by natural causes than if some other agent is holding you down. In each case, the significant thing is that you *cannot* do something you want to do. It is true that you might feel resentment against your purposeful controllers, whereas resentment is misplaced against natural forces... Now, if the distinction between CC control and mere determination by natural causes does not make a difference in our powers, when the results are the same, why should the distinction between CNC control and mere determination by natural causes make a difference in or powers, when the results are also the same? (Kane 1996, p. 68)

But secondly, I believe that there is another way to put pressure on the significance of Dennett's distinction. Ironically, the way in question involves using that favourite tool of compatibilists: the Frankfurt-style case.

I argued in the previous chapter that traditional Frankfurt-style cases do not work. However, I argued that Frankfurt-style cases cannot show an *incompatibilist* that alternative possibilities are not needed for responsibility, because the Frankfurt-style case has to assume determinism. But nothing I argued suggested that Frankfurt-style cases do not work for compatibilists. Furthermore, I did allow that Fischer's version worked to some degree – it presents some challenge to the idea that we need path-picking control. That is enough for my argument here. Note also that nearly all the compatibilists I take myself to be addressing here – so Dennett, Frankfurt, Fischer and Ravizza, Haji – all think that Frankfurt-style cases work. So my argument will have bite against them.

Consider first a standard case of global CNC control. Imagine a society of the future in which there has been sufficient scientific progress that it is now possible to 'construct' a human foetus, and to implant relevant actional elements. It is possible, given the extraordinary amount of information these engineers have, to ensure that the constructed foetus will develop into a certain kind of person. In fact, the engineers can fine-tune to

the extent that all of the agent's future actions can be ensured. In other words, if, say, they want someone to perform action X, in circumstances S at time t in twenty years time, they can engineer a foetus so that it will develop into a person who, at time t, will be in circumstances S, and will perform act X.

For many, including myself and including many compatibilists, our intuitions are that such Global CNC control subverts responsibility. However, as we have seen, the compatibilists want to say that if Global CNC control had been absent, and the foetuses had developed naturally, then free will responsibility would be in principle possible for the adults that develop from such foetuses.

But now consider that exactly the same technology that allows the brave new world engineers to ensure that their constructed foetuses develop in the right kind of persons, also allows the engineers to tell what kind of a person a naturally occurring foetus will develop into. For reasons to do with costs, these engineers prefer to deal with naturally occurring foetuses rather than constructing them from scratch. What happens is that the engineers monitor the natural foetuses, and if they see that a foetus is going to develop into the right kind of person (a person who will do all the things they want them to at appropriate points) then they do not bother intervening. Compare two adults, Larry and David, both of whom are identical in all of their character traits and other time-slice properties, but only one of whom – Larry - was subject to active intervention on the part of the genetic engineers. David had naturally been just as the controllers wanted him to be, and so intervention had not been necessary.

In this kind of case, the compatibilist would have to say that Larry is not responsible, whilst David is. Notice that in David's case he has been subject to a global version of a Frankfurt-style case. By hypothesis, the intervener in a Frankfurt-style case is *irrelevant* to an assessment of an agent's responsibility, provided the intervener does not actually intervene. But the only difference between the case in which all the foetuses are manipulated, and the case in which they are not, is that in one case we have active CNC control, whilst in the other the control is passive.

But now we can see that what explains why the agents are morally responsible when passively globally controlled but not when actively globally controlled cannot now be

being part of someone else's project. Both Larry and David are as much part of someone else's project. David may set his own ends, but the controllers allowed this only because David's ends accorded with what they wanted for David. In this respect there is no relevant difference here between passive and active covert control. After all, if the covert control is active it is still the case that you set your own ends in the compatibilist sense, for the controllers will be setting their ends via your control mechanism. So, whatever the relevant difference is between active and passive covert control which explains why (from the compatibilist's lights) David is responsible whilst Larry is not, it cannot plausibly be anything to do with being part of someone else's project. So, we should reject Dennett's explanation. Frankfurt-style cases strongly suggest that it is irrelevant that one is part of someone else's project. If active global covert control undermines moral responsibility it must do so for some other reason.

If we return to the difference between active and passive global cover control I suggest that the most obvious explanation of why the active variety undermines moral responsibility whilst the passive does not is that active involves causal determination of the agent's will. In other words, what jeopardises ownership is being determined, whether by nature or covert controller. So Frankfurt-style cases, far from supporting compatibilism, can actually be used to highlight the problem of CNC control.

Here is another way to make the same point. In a standard Frankfurt-style case the counterfactual intervener models one aspect of determinism. The counterfactual intervener, as with determinism, rules out alternative possibilities. Now, the counterfactual intervener passively covertly controls the agent in the actual sequence, but by hypothesis their presence is irrelevant to the agent's moral responsibility. However, if this intervener had intervened, their presence would be relevant to the agent's moral responsibility – they would undermine the agent's moral responsibility. But what has changed? Well, rather than passive covert control we have active covert control. So, passive covert control does not undermine moral responsibility, but active does. The relevant difference between passive and active CNC control is that one involves the determination of the agent's will whilst the other does not.

My point above then is that just as the counterfactual controller in a Frankfurt-style case *models* one aspect of determinism – namely the fact that it extinguishes our alternative

possibilities - the active controller be seen as modelling the other aspect of determinism, namely antecedent determination.

4.7 Hard Compatibilism

My arguments in this chapter will work only on those compatibilists who accept that being subject to CNC control undermines one's free will and moral responsibility. This is most compatibilists, however, there are exceptions. Hobbes is probably the most famous compatibilist who held the hard line. Hobbes believed that God controls all that we do, and that this does not threaten our free will. As Kane puts it on Hobbes's behalf, "[s]ince we do not know what God has predestined us to do, we must go on deliberating about the best way to live anyway, just as if God had not predestined us... [t]he only freedom we can have, Hobbes says, is the freedom to do what we will, and this freedom we can have even if we are predestined" (Kane 1996, p. 67). But there have been more recent advocates of this bite-the-bullet approach.

Ferdinand Schoeman is one such compatibilist. Regarding Global manipulation Schoeman holds that the agent who has had their entire reason-responsive mechanism artificially implanted is still morally responsible for their subsequent decisions and choices. He says:

> I have been maintaining that even in the event of global manipulation, the person altered is every bit as responsible for his subsequent behaviour as is the model on which he is moulded... So long as a person has the capacity to think and act on the basis of relevant reasons, he is responsible for his subsequent behaviour. (Schoeman 1978: p. 296)

Schoeman's argument is that there is no "criterion that distinguishes induced from natural desires that is relevant to the responsibility/non responsibility issue" (1978, p. 295). Because he starts out as a compatibilist about moral responsibility he sees the lack of an adequate criterion as implying that global manipulation must be compatible with moral responsibility.

Despite struggling with the issue of CNC control, Harry Frankfurt also concedes that if a

manipulator succeeds in providing a person with a new character, then "[t]hat person is then morally responsible for the choices and the conduct to which having the character leads" :

> We are inevitably fashioned and sustained, after all, by circumstances over which we have no control. The causes to which we are subject may also change us radically, without thereby bringing it about that we are not morally responsible agents. It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberate manipulative designs of other human agents. (Frankfurt 2002)

I agree with Schoeman and Frankfurt that there is no "adequate criterion" of ownership that would distinguish between determinism and covert control. That is precisely what I have been arguing here, and what I have claimed Frankfurt-style cases help highlight. But, as is clear, I see this as implying that determinism undermines moral responsibility as surely as covert control does. If a compatibilist just wants to dig their heels in and insist that global manipulation is not a problem then there probably is not much that I can say to them. There isn't anything I'd want to say – for as far as I am concerned it is a boundary condition on an adequate theory of free will that it rules out responsibility in manipulation cases. I believe that a position on free will and responsibility that fails to fulfil this condition, is going to have an extremely hard time selling itself to any not already fully committed to it.

4.8 Conclusion

In the last chapter I argued that Frankfurt-style cases do not constitute a successful challenge to PAP. However, I allowed that Fischer's version cast some small doubt on whether path-picking control is needed for responsibility. Whilst that does take some of the heat out of the luck-objection, I have argued in this chapter that compatibilism faces another challenge. Namely, it has to show that it can provide an adequate account of 'ownership'. In common with others I have argued above that cases involving CNC manipulation subvert responsibility and strongly implicate determinism as the real culprit. I have also brought to attention the way in which Frankfurt-style cases themselves can be turned against the compatibilists and help implicate determinism as

the root problem.

The arguments I have made suggest that the real problem with determinism may not be that it deprives us of path-picking control (and thereby exposes us to ultimate luck), but could be more to do with depriving us of ownership over what we do – the kind of ownership that we see CNC control depriving us of. This is just an interim conclusion. As I pointed out in the previous chapter, Fischer's Frankfurt-style case is hardly decisive, and so we are a long way short of being able to dismiss the requirement for path-picking control. It may be that indeterminism cannot provide the kind of protection against CNC control required for ownership. What we might say is that at this point we do have good reason *not* to be compatibilists, but we do not yet know whether incompatibilism fares any better. It is to incompatibilism that I turn in the coming chapters.

Chapter 5

Indeterminism and Control: I

In the previous three chapters my concern was with compatibilism and the implications of determinism for free will. I argued that there were two principle concerns about determinism. Firstly, if determinism is true, then although we can still be said to exercise control over what we do, ultimately everything we do turns out to be a matter of luck – our circumstantial and constitutive luck. This fuels the thought that what we need for responsibility is something more than compatibilists control: we need a type of control that can clear some kind of luck-free foothold. We can call this the luck objection against compatibilism. The second concern was to do with having independence from the past and ownership over what we do. Cases involving CNC control drew attention to the inadequacies of compatibilist conceptions of ownership. If we are subject to CNC control, then there is a very real sense in which our acts are not truly *ours*, and there is no relevant distinction between CNC control and determinism.

But are matters any better with indeterminism? This is the question to which I turn in this, and following chapters. My answer is a qualified yes. Briefly my answer develops over the following four chapters in the following way. In this chapter and the next my concern will be with control rather than ownership. I will argue that indeterminism does nothing to enhance our control, but nor does it necessarily damage it either. This means that in terms of exposure to ultimate luck, we are no less exposed than if determinism were true. In Chapter 7 I will turn my attention to the ownership condition that we saw compatibilists had difficulty meeting in Chapter 4. I will argue that indeterminism can provide us with the kind of ownership necessary for moral responsibility. I will consider various criticisms and will make refinements as necessary. Worries about luck persist though, and in Chapter 8 I will try to alleviate such concerns. I will point out that path-picking control, even if we had it, would do little to limit our exposure to luck, and that the implications of consistent anti-luckism are radical in the extreme. I will argue that there is a more plausible rationale behind the supposed need for path-picking control which, when recognised, goes a long way towards showing why it might not actually be necessary after all.

Turning now to this chapter. As noted above, in this chapter I argue that indeterminism does not enhance our control. I make the familiar charge that path-picking control is of questionable coherence and can be had only at the cost of mystery. I dismiss extra factor incompatibilists in the first part of the chapter. In the second part of the chapter I consider modest 'event causal' incompatibilist positions, focussing mainly on Robert Kane's variation. I argue that these positions, though they do their best to show in a non-mysterious way how our control can be enhanced by indeterminism, ultimately fail to make a convincing case. In the final part of the chapter I consider Kane's careful responses to the luck objection, and dismiss each in turn. I argue that modest incompatibilists invite confusion when they talk about 'ultimate' control, when what they really mean is exclusive compatibilist control. Kane is as guilty of this as anyone else, and as we shall see in coming chapters, it actually obscures part of the positions merit.

5.1 The Luck Objection

Incompatibilists hold that free will and responsibility require, at a minimum, the *falsity* of determinism. For only if determinism is false is it possible for an agent to have genuinely available alternative possibilities, and so only if determinism is false is there any prospect of the agent having control of a kind that would prevent everything being ultimately a matter of luck, and only if determinism is false could the agent be the originator of their actions.

Let us say that an agent has genuine alternative possibilities if the agent, S, at time t, could decide to X, or could decide to Y, consistent with the conjunction of all the facts of the past relative to t, and the laws of nature. Let us imagine that in the actual world Jones decides to X at t, and that Jones had genuine alternative possibilities. That means that if we rewind the clock to just before t, and then play things through an indefinite number of times then in some re-runs Jones will decide to X, but in some he will decide to Y. Note that everything is being kept fixed just prior to t, including Jones's values, beliefs, and motivational states.

These cases – cases where the clock is rewound and the sequence run-through again – are called roll-back cases. [59] They form the basis of the well known luck-objection to incompatibilism, for after all "if there is nothing about the agents' powers, capacities, states of mind, moral character and the like that explains this different in outcome... the difference is just a matter of luck" (Mele 1998, pp. 582-583). [60] Following Kane we can lay out the luck objection slightly more formally as follows:

> a) In the actual world, person *P*... does A at t.
>
> On the assumption that the act is undetermined at t, we may imagine that:
> b) In a nearby-possible world which is the same as the actual world up to t, *P** *(P*'s counterpart with the same past) does otherwise (does B) at t.
> c) But then (since their pasts are the same), there is nothing about the agents' powers, capacities, states of mind, characters, dispositions, motives, and so on prior to t which explains the difference in choices in the two possible worlds.
> d) It is therefore a matter of luck or chance that *P* does A and *P** does B at t.
> e) *P* is therefore not responsible (praiseworthy or blameworthy, as the case may be) for A at t (and presumably *P** is also not responsible for B) (2003, pp. 310-311)

What the luck objection does is challenge the control incompatibilist to explain exactly how indeterminism can help in terms of control – how it can offer us any enhancement of our control over what we would have if determinism were true? But that is just one aspect of the challenge presented by the luck objection. The luck objection is sometimes also used to support the claim that indeterminism would actually work to make matters worse in terms of control. In other words, indeterminism not only fails to help, it does positive damage too. One would have more control under determinism.

I think it is useful to keep these two versions of the luck objection apart and to deal with them separately. I will call the first version of the luck objection the enhancement

---

[59] What I am about to call the luck objection is also often referred to as the 'Roll Back' argument. van Inwagen first discussed the roll back argument in his 1983 (p. 141). See also Fischer 1999a, pp. 100-103 and Mele 1995 pp. 195-209.

[60] It has been made or discussed in one form or another by all of the following: Balaguar 2004; Bernstein 1995; Berofsky 2000; Clarke 1995 & 2002; Double 1991 & 1996; Fischer 1999a & 1999b; Haji 1999a;

argument, because it challenges the incompatibilist to show how indeterminism can possibly enhance our control, and I will call the second version the erosion argument, because it claims that indeterminism diminishes our control. In this chapter my focus will be the enhancement argument, with the next being devoted to the erosion argument.

Before we consider the enhancement luck objection we can note that roll back cases do highlight that, if indeterminism is true in such a way as to yield genuine alternative possibilities in which the agent decides otherwise, then not everything an agent does is ultimately a matter of their constitutive and circumstantial luck.[61] Clearly P and P's counterpart P* have the same constitutive and circumstantial luck, yet make different decisions. So a different kind of luck has to be invoked here. From now on, this is how I will be using the term causal luck. We can say that according to the luck objection, if indeterminism is true in such a way as to yield genuine alternative possibilities, then everything an agent does is a matter of their constitutive, circumstantial and *causal* luck, whereas if determinism is true, then everything an agent does would be a matter only of the former two.

Although I have been using the term in previous chapters, let me make clear that in what follows I will refer to those who maintain that in addition to having genuine alternative possibilities an agent needs 'path-picking' control as control incompatibilists, and I will use path-picking control and incompatibilist control interchangeably.

With respect to the luck objection, the control incompatibilist is going to have to insist that if the agent exercises incompatibilist control over their decision, then it was no matter of luck that the agent made one decision rather than another. So, whereas just straightforward indeterminism would (we are assuming) render Jones's decision a matter of causal luck, the introduction of incompatibilist control eradicates such causal luck. It would have been a matter of causal luck which way Jones decided, but with incompatibilist control in the picture, it becomes a matter over which Jones exercises control.

2002; Kane 2002; 2003; Mele 1998; 1999; 1999b; O'Connor 2000; Strawson 2000 & 2003; Smilansky 2000; Waller 1988.

[61] Here I contradict something that Latus (2001) has said. Latus claims that of Nagel's categories of luck, 'causal' luck is redundant for it is fully captured by circumstantial and constitutive luck.

But there are difficulties when it comes to understanding what incompatibilist-control might amount to, and how we could possibly have it.

In what follows I will argue that unless an appeal to mystery is made, there is currently no clear way to see how indeterminism, wherever it is located, could enhance an agent's control in such a way as to clear a luck-free foothold.

5.2 Path-Picking Control

What must path-picking control involve? This is an extremely difficult question to answer but as a first attempt we might say that it involves having the ability to antecedently ensure which pathway one travels down. Following Kane we can term this control Antecedent Determining Control:

> the ability to be in, or bring about, conditions such that one can guarantee or determine which of a set of outcomes is going to occur *before* it occurs, whether the outcomes are one's own actions, the actions of others, or events in the world generally. (Kane 1996, p. 144)

This is a kind of control that we value in everyday life and that we try to attain. When we pick up a particular skill we have acquired an ability to ensure, within certain limits, that certain things happen. My skill at archery involves my ability to ensure that most of my arrows hit their target. We can note that this is a kind of control that is compatible with determinism: the truth of determinism would not lead us to doubt that we have these kinds of abilities, and that we can exercise such abilities when we want. Compatibilist control just is antecedent determining control of some degree. This much we have seen in Chapter 2.

But what the control-incompatibilist needs to do to get genuine path-picking control of a kind that could eradicate causal luck is fuse possession of antecedent determining control with genuinely available alternative possibilities. But this looks impossible. The problem is that our decisions could only be things over which we exercise antecedent ensuring control if we have a deeper self exercising such control. But the same problem arises at the level of the deeper self. For presumably this deeper self has a will too.

Either its will operates deterministically, or indeterministically. If determined, then it turns out that all that we do ultimately turns out to be a matter of circumstantial and constitutive luck operating at the level of the deeper self. If the agent's deeper self operates indeterministically, it looks as if we have causal luck and nothing has been said to meet the original luck argument. It would be no good invoking a yet deeper self, for exactly the same problems would arise for that self too. Susan Wolf is one amongst many who has made this point:

> In order for an agent to be autonomous, it seems, not only must the agent's behaviour be governable by her self, her self must in turn be governable by her self – her deeper self, if you like – and this must in turn be governable by her (still deeper?) self, *ad infinitum.* If there are forces behind the agent, so to speak, making the agent what she is, then her control of her behaviour is only intermediate, and therefore superficial. But if there are no forces behind the agent making the agent what she is, then her identity seems to be arbitrary. (Wolf 1990,p. 14)

Some might think the invocation of a soul or a self in some immaterial realm might be in order. But none of these extravagant strategies will work either.

> I would think that the 'self', even if immaterial, would act because of its particular attributes and that those attributes are inherited and undergo development with experience. We would have no reasonable explanation for how the self gets to be the way it is unless it derives its character potential from birth. We could not account for consistent behaviour on a person's part unless the self has an enduring structure. (Walter 1978, p. 509, quoted in Smilansky 2000, pp. 64-65)

Thus neither invoking deeper selves, nor immaterial selves occupying some other realm will be of any help in addressing the basic problem. But perhaps this is because we are trying to understand path-picking control in terms of compatibilist control. What, then, if we try to understand incompatibilist control as an altogether different kind of control? Namely a kind of control that isn't a function of the way the agent is at a particular time – that isn't to be understood as some process unfolding in the agent. The problem

though, is understanding exactly what this different kind of control could be, and how it could eradicate causal luck. Although it fell out of favour for a while, a number of contemporary incompatibilists talk about a distinct kind of causation – agent-causation. This kind of causation is causation by an agent – a thinking or rational substance. As van Inwagen puts it:

> The friends of agent causation hold that the causes of some events are not (or are only partially) earlier events. They are rather substances – not *changes* in substances, which are of course events, but "substances themselves". Thus, they say, Thomas Reid caused the movements of his fingers when he wrote the sentence, "There is no greater impediment to the advancement of knowledge than the ambiguity of words." These movements, they insist, were caused simply by *Reid*, and not by any change in Reid. (van Inwagen 2002, p. 169)

But again, in common with many others I see the words agent causation as nothing more than a label for a mystery. Maybe it can make sense, maybe there is the distinct type of causation in question, but it is mysterious how exactly this provides path-picking control apart from merely stipulating that it does. Possibly some agent-causalists hold that agent causation makes an agent morally responsible just because it is now true that the agent caused the decision that they made, and that this happening was not determined to happen by any prior state of affairs. But this does nothing to show how agent-causation provides anything in terms of control over what one would have in a deterministic setting, or how the agent in the roll-back case can now be said not to be subject to causal luck. Van Inwagen makes this point nicely below:

> [In roll-back scenarios the agent-causalist] might say this: If it turns out that Alice agent-causes truth-antecedent cerebral events, this will not be a matter of chance because it will be she, *Alice,* who is the cause of the event "its coming to pass that Alice agent-causes truth antecedent cerebral events." But have we not got every reason to regard the occurrence of *this* event – that is, the occurrence of "its coming to pass that Alice agent-cause the event 'its coming to pass that Alice agent-causes truth antecedent cerebral events'" – as a matter of chance? If the three events "the truth-antecedent cerebral events"/"its coming to pass that Alice agent-causes the truth antecedent cerebral events"/

> "its coming to pass that Alice agent-causes the event 'its coming to pass that Alice agent-causes truth –antecedent cerebral events'" are the first three terms of an infinite series of agent-caused events, is not the simultaneous occurrence of all the events in this sequence (as opposed to the simultaneous occurrence of all the events in an infinite sequence of agent-caused events whose first member is "lie-antecedent cerebral events") a mere matter of chance? (2002, p. 174)

In common with many others, I hold that incompatibilist-control is incoherent when understood in terms of compatibilist antecedent ensuring control, and ineliminably mysterious when understood in terms of agent causation.

5.3 Modest Incompatibilism: Valerianism

I now want to turn to incompatibilists who have tried to rise to the challenge of showing, in a non-mysterious way, just how the twin demands of alternative possibilities and control can be met. Such positions do not make recourse to special kinds of causation, but instead try to get by with just straightforward event causation. Following Haji I will refer to these, honest, down-to-earth brands of libertarianism as modest libertarian positions.

For our purposes we can distinguish two kinds of modest libertarian position.[62] The difference concerns the location of the indeterminacy, and the first kind that we are going to consider is often called Valerian incompatibilism.[63]

To see what Valerian incompatibilism amounts to, start with a reason-responsive mechanism of the kind that a compatibilist can agree delivers the sort of control

---

[62] I have chosen to ignore non-causal incompatibilist views of the type most prominently defended by Carl Ginet (1990; see also Goetz 1997 and McCann 1998). According to this view free actions have uncaused simply mental actions at their core. What marks out these mental events from other mental events is their "actish phenomenal quality" (Ginet 1990, p. 13). However, by denying that these mental events are caused this view faces the problem of explaining in what possible sense they are controlled at all (Mele 1992; O' Connor 2000; Clarke 2002). What I argue here and in coming chapters is not jeopardised by having ignored non-causal views. Indeed, to some extent my arguments - because they ultimately involve placing the stress on considerations to do with ownership rather than control – help non-causal accounts. So I believe it to be safe to ignore such views, and warranted on grounds of clarity and space.

[63] The term was first used by Bernstein because Dennett (who I think was the first to suggest this kind of incompatibilism) cited the poet Valery (Dennett 1978; Bernstein 1989; see also Kane 1996, p. 236 n. 9).

sufficient for moral responsibility. So, the agent has the necessary psychological basis for evaluative reasoning, which will include beliefs, values, and desires, and a background evaluative scheme. However, during the process of deliberation it is causally open which beliefs will come to mind. This does not mean that the output – the decision – that the mechanism delivers as a result of this process is uncontrolled. It just means that it is now indeterministic how the process will turn out, because it is causally open what ingredients – in the form of beliefs – will be fed in.[64]

Because it is open what beliefs will come to mind, introducing indeterminism in this way will yield genuinely alternative possibilities in which the agent decides otherwise.[65] A major advantage of creating the agent-internal indeterminism by making it indeterministic which beliefs come to mind is that the indeterminism does not erode compatibilist control. In terms of compatibilist control the resulting decision is as controlled as it would be were the entire process deterministic. After all, we do not have control over which beliefs come to mind irrespective of whether we are in a deterministic or indeterministic universe. As Haji puts it, this view provides "indeterministic agency while impeding or restricting our control over what happens only in domains in which we have no greater control on the hypothesis that our world is deterministic... even if determinism is true, it is false that, with respect to each consideration – belief, desire , and so on – that comes to mind during our deliberation, we are in control of its coming to mind; and some considerations that come to mind without our being in control of their doing so doing may influence the outcome of our deliberation" (Haji 1998, p. 28 & see also Mele 1995 ch. 12). Mele explains how this contributes to our control:

> Considerations that indeterministically come to mind (like considerations that

[64] Alfred Mele, is one who has developed the view in question (though he does not fully endorse it – Mele is agnostic between compatibilism and modest libertarianism) (1995, p. 215). Dennett has also suggested a version of valerian incompatibilism, but only for the purposes of rejection, and likewise with Fischer (Dennett 1978; Fischer 1995). Some have also taken Ekstrom's view to be Valerian (see Clarke 2002). However, Ekstrom thinks that free will requires that an agent's preference formation be indeterministic, but for Ekstrom decision making is a form of preference formation (Ekstrom 2000, p. 107). This is something that Balaguar has pointed out, and he says that Ekstrom has confirmed it in private correspondence (see Balaguar 2004, p. 380 n. 2).

[65] It will not necessarily be the case that the agent could have decided otherwise. It may be that it was indeterministic whether a particular belief would come-to-mind during an episode of deliberation, yet irrespective of whether this belief would come to mind or not it was still inevitable that the agent would decide as they did. Needless to say, if this were *always* the case then the agent in question would never have the kind of free will necessary for moral responsibility – at least not by this view's lights.

> deterministically come to mind) are nothing more than input to deliberation. Their coming to mind has at most an indirect effect on what the agent decides, an effect that is mediated by the agent's own assessment of them. They do not settle matters. Moreover, not only do agents have the opportunity to assess these considerations, they also have the opportunity to search for additional considerations before they decide, thereby increasing the probability that other relevant considerations will indeterministically come to mind. They have the opportunity to cancel or attenuate the effects of bad luck (for example, the undetermined coming to mind of a misleading considerations or an undetermined failure to notice a relevant consideration). And given a suitable indeterminism regarding what comes to mind in an assessment process, it is not causally determined what assessment the agent will reach. (Mele 2002, pp. 544-545)

However, the kind of antecedent ensuring control that one has here is exactly the same kind as one would have were it deterministic which beliefs would come to mind. One does not have luck-eradicating control over how one responds to the beliefs that come to mind. The agent's response will exhibit control, insofar as it will be the output of a reason responsive mechanism, but this will do nothing to prevent everything the agent decides being ultimately a matter of luck. For it remains the case that Jones reacts in the way that he does to the beliefs that do come to mind because of the way that he is. Fischer makes the same point:

> How can adding arbitrariness of the sort envisaged – the lack of determination of the beliefs that come to mind during deliberation – to a causally deterministic process yield genuine control? A libertarian... will contend that an *entirely* deterministic process does not contain genuine control by the relevant agent. How, then, can installing the sort of indeterminacy envisaged – indeterminacy as to which belief states will come to the agent's mind – transform the sequence from one of lack of control to one containing control? This smacks of alchemy. (1999b, p. 140)

But it might be objected that indeterminism introduced in this way *does* enhance an agent's control insofar as it delivers a species of *ultimate* control. For with such

indeterminism in place then the agent gains a special kind of "agency that gives them a kind of independence and an associated kind of explanatory bearing on their conduct that they would lack in any deterministic world" and they would "make choices and perform actions that lack deterministic causes in the distant past" (Mele 2002, p. 545). Mele sees the indeterminism as providing ultimate control, where this is understood as an agent making a decision for which there were not causally sufficient prior conditions that were external to the agent (Mele 1995, p. 211).

But I think only confusion comes from conceiving of this as an enhancement of control, though as we shall see the habit of doing is common to nearly all modest incompatibilists. What ultimate control actually means is that the agent achieves a kind of exclusive control (though I think it is questionable whether Valerian views can provide sufficient exclusivity – a point that I will be making in greater detail in Chapter 7). Such indeterminacy might be seen to answer the first concern about determinism, namely that if determinism is true then we are just links in a chain and so are not the true sources of our actions. (As will become clear later, I question whether the view can do even this). But nevertheless, those are the kinds of concern that this view must be seen as addressing. What it does not do is enhance an agent's control in a way that would answer the luck objection. For the kind of exclusive control that the agent gets, is exclusive compatibilist control. Ultimate luck remains.

5.4 Modest Incompatibilism: Kane

We have seen that an internally indeterministic mechanism cannot provide the kind of path-picking control wanted if the indeterminism is located early in the deliberative process, but maybe matters change if the indeterminism is moved so that it is much later in the deliberative process. If it is indeterminate what decision the mechanism will issue at the moment of choice, then perhaps this could be said to deliver what is wanted. It is indeterminism up to the moment of choice that defines a non-valerian incompatibilist. I will use the term modest incompatibilist to refer to non-valerian incompatibilists from now on.

Clarke refers to incompatibilist positions which insist upon making the agent's decision non-deterministically caused "action-centred", because making a decision I a mental

action (2002). Robert Kane is the most prominent of these modest incompatibilists, though others who have outlined such views include Wiggins (2003); Sorabji (1980); Nozick (1981, pp. 294-316); Balaguar (1999 and 2004); and Ekstrom (2000). Kane though has done most to develop a detailed account of this kind of view in his work *The Significance of Free Will* (1996). It will be his view that will provide my primary focus in the forthcoming discussion, though other similar incompatibilists will also be referred to. I will begin by giving an all too rough outline of this view, but one that will serve well enough for our purposes here.

On Kane's account our free will stems from occasions of motivational conflict, where an agent is torn between different courses of action – torn decisions in other words. Kane calls the decisions that resolve such internal conflicts Self Forming Acts or SFAs, and he suggests that there are six kinds (though he does not claim these to be exhaustive:

> [Self Forming Acts] include acts of the following kinds: 1) Moral choices or decisions, 2) prudential choices or decisions, 3)efforts of will sustaining purposes, 4) attentional efforts directed at self control and self modification, 5) practical judgements and choices, and 6) chances of intention in action. (1996, p. 125)

1 and 2 involve "conflicts between what an agent believes ought to be done and what the agent wants or desires to do" (Kane 1996, p 126). In the case of prudential conflict this involves being torn between doing what is in one's long term interests, and what is in one's immediate interests. In the moral case the conflict is between what one takes to be morally required, and what one wants to do. One has a practical conflict where one is torn between options neither of which has moral or prudential reasons in its favour (so Buridan's ass cases would be cases of practical conflict as would be conflicts between values).

So, on the occasion of a torn decision the agent recognises reasons for two options, and they are torn about which way to go, and neither set of reasons seem to outweigh the others. The agent's prior character and motives provide both the reasons why the agent is trying to do, say, their duty, *and* the reasons why the agent is trying to do otherwise.

As Kane puts it, the agent's "complex of past motives and character... explain the *conflict* within the agent's will *from both sides*" (1996, p. 127). Ultimately the agent ultimately just chooses. It is this kind of case that Kane asks us to imagine could be indeterministic so that the "might choose either way, all past circumstances remaining the same up to he moment of choice" (1996, p. 127).[66] Kane offers us an analogy:

> Consider a quantum analogue. Imagine an isolated particle, such as an electron, moving toward a thin atomic barrier. Whether or not the particle will penetrate the barrier is undermined. There is a probability that it will penetrate, but not a certainty, because its position and momentum are not both determinate as it moves toward the barrier. Imagine that the choice (to overcome temptation) is like the penetration event. The choice one way or the other is *undetermined* because the process preceding it... (i.e., the effort of will to overcome temptation) is *indeterminate*. (1996, p. 128)

Applying this analogy to torn decisions – or SFAs – he elaborates further:

> There is a tension and uncertainty in our minds at such times of inner conflict which are reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium – in short, a kind of stirring up of chaos in the brain that makes it sensitive to micro-indeterminacies at the neuronal level. As a result, the uncertainty and inner tension we feel at such soul-searching moments of self-formation is reflected in the indeterminacy of our neural processes themselves... When we do decide under such conditions of uncertainty, the outcome is not determined because of the preceding indeterminacy – and yet it can be willed... either way owing to the fact that in such self formation, the agents' prior wills are divided by conflicting motives. (2003, p. 306; see also 1996 chs. 8-10)

---

[66] There is no need to conceive of torn decisions as only concerning important issues such as whether to act morally or self-interestedly. As Balaguar says, "we make [torn decisions] ..all the time, every day of our lives".

> To appreciate this, consider a second case, involving a decision more mundane than Ralph's. Jane is in a restaurant, deliberating about whether to order tiramisu or a fruit plate for desert. She thinks that the former will taste better but that the latter will be better for her health. She has no clue which reason is stronger and feels genuinely torn. Suddenly, it's her turn to order; the waiter is looking at her; she has to pick; Oh, God, "I'll have the tiramisu," she says. I.e. she just chose. Period. (Balaguar 2004, p. 363)

I said above that Kane outlines 6 different types of SFAs or torn decisions. So far I have mentioned moral conflicts, prudential conflicts and practical conflicts. I will briefly outline the remaining three.[67]

5.5 Efforts Sustaining Purposes

As Kane sees it, when we resolve a torn decision the "reasons for which one chooses causally influence the choice (without determining it), while the choice, once made, reorganizes the motivational structures of the brain so that the reasons come to have a special role to play in future behaviour" (Kane 1996, p. 139). So we can understand our SFAs as 'will setting' in this respect. But sometimes we have to make an effort to sustain a purpose rather than form one. The effort arises due to there being a conflict between sustaining the already-formed purpose or intention, and one's inclinations, such as one's inclination towards laziness, or one's fears or dislikes.

5.6 Acts of Attention

Sometimes we could ensure that we make a particular decision if we could bring to mind a certain image. Kane gives an example of a man struggling to give up smoking who can only do so when he brings to mind an image of his father dying with lung cancer. But to do this itself takes effort, and this gives rise to another occasion on which we might find our will conflicted:

> [T]he focussing of attention may itself be difficult and require effort, not only because it may be difficult to concentrate, or the object of attention may be unpleasant... but also, and more generally, because there may be resistance in the will toward doing anything to temper or control what the agent otherwise

[67] There is room for more to be added. Amongst modest incompatibilists there can be disagreement over just when and how often one faces torn decisions. Van Inwagen, for instance, holds that there are only three types of SFA. We have SFAs in 'Buridan's Ass' cases (corresponding to Kane's practical SFAs) (1989, p. 405). Secondly there is the case where a duty that one takes oneself to have conflicts with one's inclinations or desires – corresponding to Kane's moral SFAs (van Inwagen 1989, p. 405). The third kind of case involves internal conflict that arises through having to choose between incommensurable values. When discussing Kane I have considered this a practical conflict (van Inwagen 1989, p. 405). Van Inwagen also holds that we *rarely* face such situations and as such we rarely perform *directly* free acts. But Kane, and another modest incompatibilist – Balaguar – both think that we face SFAs quite regularly.

> strongly wants to do... During periods when the agents are making such efforts to attend against resistance within their wills, the efforts could be conceived as indeterminate efforts. (Kane 1996, p. 157)

5.7 Changes of Intentions in Action

When an agent comes to an all-things considered best judgement about what to do, but then does otherwise then e have a case of what Mele calls strict akratic action (1987, p. 19). Assuming that such cases are possible, then such cases can also provide instances of SFAs of a particular kind. In these cases the agent has settled on a course of action, but then decides otherwise at the last moment. Kane suggests that these cases can be accommodated by his view in the following way. What goes on in such cases is a sudden build up of tension due to a suppressed conflict, rather than one preceded by deliberation as in the more usual SFAs. And "[t]he effort to resist these suddenly strengthened inclinations would then be indeterminate, like previous efforts required to overcome countervailing inclinations" (Kane 1996, p. 170).

As I said, Kane does not propose the above as an exhaustive list of all the possible kinds of occasions in which we face an internal conflict, and there is room for disagreement over the length of such a list. However, the principle focus is on moral conflict cases, so, cases where there is an internal conflict in the agent between what they consider they morally ought to do, and what they consider they prudentially ought to do. Such cases will also be my focus in the remainder of this chapter.

5.8 Kane and the Luck Objection

On first inspection one might wonder how Kane's view does not just describe, in detail, what is going on in roll-back cases. For if an agent's effort of will is indeterministic in any of the six ways he outlines, then, holding fixed the agent's motivational states etc, and past circumstances, one will, if one runs the sequence through an indefinite number of times, have occasions in which the agent decides one way, and occasions in which they decide another way. In other words, one seems to have just causal luck. How does this view do anything to show how causal luck can be eradicated and a luck-free foothold established?

In my view Kane cannot show this. I actually endorse Kane's account of free will, but unlike Kane I do not think that his account offers us anything more in terms of control than would be available under determinism. In other words, in terms of control, all this account offers is compatibilist control. But Kane himself thinks otherwise – or at least, certainly seems to. He has given over much time and ink to trying to answer the luck objection. I will consider his various responses below, all of which, I believe, fall short of showing how our control has been increased.

I will first put in the following qualifications. Although I am going to be saying more about this later, it would be as well to recognise here that Kane's primary concern is to secure the "power of agents to be the ultimate creators (or originators) and sustainers of their own ends and purposes" (Kane 1996, p. 4). In other words, Kane's main focus is that first concern about determinism – the concern about ownership and being the ultimate source of what we do. Kane has done more than anyone else in the contemporary debate to draw attention to this much overlooked condition. I am fully behind Kane on that score, as will become clear in later chapters. Nevertheless, Kane does *also* seem to want to say that his account secures greater control for an agent over what they could have in deterministic settings. It is that which I take issue with.

I said in the introduction to this chapter that we need to distinguish between the 'enhancement' version of the luck objection and the erosion version. Related to what I have just said above, it is not clear which version of the luck objection Kane is answering in his responses below (Kane is never explicit). If he is answering the second concern, then I believe that what he says has much merit. If he is attempting to answer the first concern however, I think what he says has very little merit. In truth, Kane is almost certainly trying to do a bit of both. But, at the risk of being uncharitable, and in the interests of clarity, I am going to assume that Kane is trying to answer the first concern – as we shall see shortly, some of what he says certainly suggests he thinks he is answering the first concern.

5.9 The Question Begging Charge

A common way for a control incompatibilist to respond to the luck objection is to accuse it of being question begging (see Wiggins 2003, p. 112 and for discussion Watson 2003, p. 10). For the luck objection simply assumes that in the rollback cases the reason that Jones decides one way rather than another is down to luck rather than an exercise of incompatibilist control. This is true – it *is* question begging - but only in the most mundane uninteresting way. It seems to me that the onus is on the incompatibilist to provide an explanation of how it could be that we could have causal-luck eradicating control, for unless this is forthcoming, the luck objection seems to have been met by fiat. It may be that in the end, if path-picking control is really thought to be necessary for moral responsibility, that this is as much as the control-incompatibilist can do. But then their case will appeal only to those already committed to incompatibilism.

Kane himself accepts all of this – he is against introducing extra factors to deal with the problem of luck. So, he is not asking us to just accept that we have a mysterious, unanalysable form of control. But he nevertheless does hold that the luck objection begs the question, because, as he puts it "'Chance' and 'luck' are terms of ordinary language which carry the connotation of "its being out of my control"... So using them already begs certain questions, whereas 'indeterminism' is a technical term that merely precludes *deterministic* causation (though not causation altogether)" (2003, p. 305).

This is a much milder version of the question begging accusation. As such I accept that there is something in what Kane says here, but not much if one is trying to show how indeterminism enhances control. I think it should be accepted that using the term luck with regard to roll-back cases *is* misleading given that the kind of luck we are talking about here is luck internal to the agent's will. Often when we use the term luck we do mean to attribute something we have done to a factor external to ourselves – external to our conscious willing self, that is. Causal luck of the type that we see evidenced in roll-back cases is not 'outside' the agent's control in the sense of being due to some factor external to the operation of the agent's conscious will. The luck is inside the agent's control. But there's an important difference between something being 'inside' the agent's control in the sense of being internal to the workings of the agent's mechanism

(their effort of will), and something being *under* the agent's control in a way that would eradicate luck.

Consider the television remote-control from previous chapters. On the 'Kane' remote control, when two buttons are pressed simultaneously – analogous to a conflict in the will - it becomes indeterministic which channel change will result. In this case would we not say that it is chancy which channel it will change to? The luck or chance is internal to a control mechanism, and granted, we would still say – or at least I would - that the resulting channel change was controlled by the remote-control inasmuch as it was sourced exclusively to its internal operation. There is also a sense in which it was under the mechanism's control to change to BBC1 rather than BBC2 - again inasmuch as nothing external to the mechanism could play any role in this. But the sense in which the channel change is under the control of the remote control is, I would say, consistent with it also having been a matter of luck that the mechanism changed channels to BBC1 rather than BBC2.

Analogous remarks apply to the situation in SFAs. The agent-internal indeterminism does secure a negative condition – it does mean that we cannot cite anything external to the agent's effort of will as an explanation of why the agent decided one way rather than the other. We can say that it was truly down to the agent which way they decided. But that is consistent with it having been a matter of luck which way the agent decided.

One might object that the barrier between what is inside, and what is outside the agent's will is one which transforms what would be lucky if outside, to controlled if inside. That is why there is a sense in which we can say of the remote-control that it controlled the channel change even though it was indeterministic which way it would change the channel. I grant this. But it will not help to show how agent-internal indeterminism can *enhance* control. For exactly the same points could be made by a compatibilist. They could point out that the fact that one does not control the inputs to one's control mechanism does not mean that the outputs are not controlled – they *are* controlled, just in virtue of being the output of a control mechanism. Yet this kind of point, whilst correct, does nothing to meet the luck objection against compatibilism and so similarly cannot help Kane either.

Whilst Kane has a point about the appropriateness of the word luck, this is at best a terminological point, and furthermore all of what he says would hold good for the compatibilist too. Because ultimately everything is a matter of luck if we only have compatibilist control, then nothing Kane has said above shows us how agent-internal indeterminism would change this.

There is another way in which talk of luck is slightly misleading, which Kane also picks up on. Consider that it is tempting to see luck as some kind of force, such that to say it was a matter of luck that Jones decided one way rather than the other, is to say that Jones was trying to decide, and then luck came in and settled matters. But as Kane says "this is the wrong picture":

> On the [modest incompatibilist] view just described, you cannot separate the indeterminism from the effort to overcome temptation in such a way that *first* the effort occurs *followed* by chance or luck (or vice versa). One must think of the effort and the indeterminism as fused; the effort *is* indeterminate and the indeterminism is a property of the effort, not something separate that occurs after or before the effort... There is no point at which the effort stops and chance "takes over". She chooses *as a result of* the effort, even though she might have failed because of the indeterminism. (Kane 2003, pp. 313-314)

Again, I think that Kane is right about this, but that it does not help. For Kane's point is analogous to one that compatibilists make about determinism. Namely that determinism is often misleadingly thought of as a force pushing an agent.[68] As Fischer points out, "[t]here is a commonsense notion of "pushing," according to which there is a difference between (say) being pushed by a strong gust of wind and simply walking down a trail" (2003, p. 209). This difference is one that the truth of determinism would not jeopardise. Similarly, there is a commonsense notion of luck according to which one fails to succeed in doing what one was trying to do due to some external factor (a strong gust of wind). This is *not* the case where decisions themselves are concerned. For in the case of torn decisions there is nothing external affecting the outcome of our effort.

---

[68] This is how Ekstrom characterises determinism.

> Jones's subjective perception of available options is irrelevant; in fact, the past pushes him into one particular decision state, the only state physically possible at the time, given the past and the

But just as the compatibilist's point does nothing to save them from the luck objection, so too Kane's point – although correct - does nothing to save him from the luck objection either. To the compatibilist one can simply reply that though it might be misleading to think of determinism as pushing an agent into making one decision rather than another, it nevertheless remains the case that it was inevitable that the agent would make the decision that he did and that as such it was just the agent's good/bad luck that he did. Likewise to Kane one can say that though the indeterminacy of the agent's effort does not interject or come-in from the outside, it remains the case that it was a matter of the agent's luck that he decided one way rather than another. All Kane has done is warn us against mistaking this claim – the claim that it was a matter of luck that the agent decided one way rather than another – as the claim that some outside force luck played a role in what happened. But he still leaves us none the wiser how ultimate luck has been eradicated. In short, I do not think that there is any room to wriggle out of the luck objection by talking about how we might usually use terms like luck.

Despite what I have just said, it is nevertheless significant that our talk of luck does not seem quite appropriate when it comes to decision-making. For it may be that talk of luck *sounds* toxic to responsibility, precisely because we associate luck with being subject to external factors. In other words, the toxicity of luck to responsibility could be owed to the fact that we think we cannot be responsible for what we do if it is sourced exclusively to factors external to our will – if nothing is left that is ultimately *down to us*. As such it may be that agent-internal causal luck is nowhere near as obviously toxic to responsibility as external luck is. I am going to say more about this in Chapter 8, for I think that it is importantly correct. Nevertheless these comments do nothing to show how agent internal indeterminism enhances control. So despite my sympathy with some of what Kane says above, we are none the wiser how control has been increased.

5.10 The 'Exact Sameness' Argument

Another argument that Kane sometimes makes concerns the assumption of prior sameness made in the roll-back cases normally used to generate the luck objection.

---

laws of nature... Jones should not be judged as morally responsible for his decision and his act, given the pushing feature of determinism. (1998 pp. 284-285)

Consider that for Kane indeterminacy is internal to the process – the effort of will – that culminates in decision-making: "the indeterminism and the effort are "fused": the indeterminacy is a property of the effort and the effort *is* indeterminate" (Kane 1996, p. 151).[69] Yet in roll-back cases everything is held fixed just prior to decision-making. Kane now objects that one cannot hold matters fixed, where the indeterminacy is internal to the agent's will.

> With indeterminate efforts, exact sameness is not defined. Nor is exact difference either. If the efforts are indeterminate, one cannot say the efforts had exactly the same strength, or that one was exactly greater or less great than the other. That is what indeterminacy amounts to. So one cannot say of two agents that they had exactly the same pasts and made exactly the same effort and one got lucky while the other did not. Nor can one imagine the same agent in two possible worlds with exactly the same pasts making exactly the same effort and getting lucky in one world and not the other. Exact sameness (or difference) of possible worlds is not defined if the worlds contain indeterminate efforts or indeterminate events of any kinds. (1996, pp. 171-172)

It is not clear why there cannot be exact sameness.[70] It is even less clear how exactly this argument is supposed to do anything to meet the luck objection. Even if Kane is right above, one could simply rewind the clock to a moment just prior to the beginning of the effort of will and run the roll-back cases from this point. One would still get two agents, identical *up to the moment of the effort of will*, and who subsequently make different choices (Clarke 2002, p. 372; Haji 1999b, p. 53; Mele 1999, pp. 279-280). Unless I am

---

[69] Kane admits, "[t]o fully understand how this fusion could take place would be... to understand the nature of conscious experience and its unity... as well as to understand how consciousness and mind are related, if at all, to the indeterminacy of natural processes... [I]n other words, it is possible that the ultimate understanding of this fusion may lie in the *connection*... between consciousness and quantum reality" (1996, p. 151). So, we have an admission of mystery into Kane's account. But the mystery in question is a mystery for everyone and not one peculiar to an incompatibilist view.

[70] Clarke certainly thinks there can:

> [I]n a straightforward sense, there can be exact sameness of one world to another even if there is indeterminacy. In physics, the indeterminate position of a particle may be characterised by a wave function (one specifying the probabilities of the particle's being fond, upon observation, in various determinate positions), and the particle (or its counterpart) in a different world may be correctly characterised by exactly the same wave function. (2002, p. 371)

missing something, this denial of exact sameness or difference does nothing whatsoever to meet the luck objection.[71]

5.11 Choosing for Reasons

Kane regularly attempts to deflect concerns about luck in roll-back cases by pointing out that roll-back cases do not imply "that [Jones and Jones*] 1) did not *choose* at all, nor does it imply that they did not both choose 2) *as a result of their efforts*, nor that they did not choose 3) *for reasons* (different reasons) that 4) they most wanted to choose for *when* they chose, nor that they did not choose for those reasons 5) *knowingly* and 6)*on purpose* when they chose, and hence 7) *rationally,* 8) *voluntarily*, and 9) *intentionally*" (Kane 2003, p. 320). Yet, as he goes on to say, "these are precisely the kinds of conditions we look for when deciding whether or not persons are responsible" (2003, p. 320).

But note first, that Kane cannot, if he wants to preserve his incompatibilism, insist that considerations 1-9 are the only ones relevant to establishing whether someone is responsible or not. For nothing about the thesis of determinism precludes any of 1-9 being met. Nothing about the thesis of determinism implies that we do not choose rationally, on purpose, as a result of our efforts and so forth. Furthermore, if determinism is true, then ultimately everything we do is a matter of luck. At the moment our concern is with that version of the luck objection which asks how agent-internal indeterminism could do anything to prevent everything we do being a matter of luck. So, simply listing a number of considerations all of which are consistent with everything we do being ultimately a matter of luck cannot therefore show us how our control has been enhanced in any way.[72]

5.12 The Resultant Luck Argument

Sometimes Kane actually tries to use a specific kind of moral luck to provide support for the compatibility of agent-internal indeterminism and responsibility:

---

[71] In fairness Kane now seems to acknowledge this (See his 2003, p. 310 fn. 18).

> Consider an assassin who is trying to kill the prime minister but might miss because of some undetermined events in his nervous system which might lead to a jerking or wavering of his arm. If he does hit his target, can he be held responsible? The answer (as J. L. Austin and Philippa Foot successfully argued decades ago) is "yes," because he intentionally and voluntarily succeeded in doing what he was *trying* to do – kill the prime minister [Austin 1961, pp. 153 – 180 & Foot 1966]. Yet his killing the prime minister was undetermined. We might even say in a sense that he got lucky in killing the prime minister, when he could have failed. But it does not follow, if he succeeds, that killing the prime minister was not his action, not something he did; nor does it follow... that he was not responsible for killing the prime minister. Indeed, if anything is clear, it is that he both killed the prime minister and was responsible for doing so. (2003, p. 308 one note omitted; see also his 1996 pp. 54-56 & 2002, p. 418)

It is not entirely clear how Kane intends that we should take his argument above. Kane's acknowledgement that we might say of the assassin that "he got lucky in killing the prime minister" suggests that he is trying to meet the second luck objection. Taken as such the argument can be interpreted as follows. It was lucky that the assassin killed the prime-minister rather than missing. It is assumed that this kind of luck (Nagel's 'resultant' luck) does not subvert responsibility. Why doesn't it subvert responsibility? Well, because the agent did what he did on purpose, intentionally, and so forth. Working backwards, we note that the same holds true at the level of decision making. In one sense it was a matter of luck which way the agent decided, but this does not work to provide the agent with an excuse. The agent still decided intentionally, on purpose and so forth. Therefore we have no reason to think that the agent did not control their decision in the appropriate way.

Interpreted in this way, it is actually an argument designed to show that luck is *not a problem*, rather than an argument designed to show how ultimate luck has been eradicated.[73] Taken as such I believe it has some merit so long as it is located in the

---

[72] It is when Kane makes arguments such as the one above that one is tempted to see him as trying to answer the 'erosion' argument.

[73] Note that this kind of argument works equally well for the compatibilist too. After all, a compatibilist could point out that removing the undetermined events so that the process from decision to the prime-minister being shot is deterministic, does not undermine responsibility either. The compatibilist could then

context of a wider argument designed to show that luck is not a problem – an argument that I will be making in Chapter 8. However, taken in isolation, the argument is not convincing. For the fact is that our intuitions about resultant luck are not clear. Of all the kinds of luck, it is the one we are most likely to be troubled by. This means that Kane's argument here lends itself to being used to motivate the thought that causal luck *is* a problem. After all, plenty agree that on reflection the assassin *is not* responsible for having killed the prime minister, precisely because this was a matter of luck (Sverdlick 1993, p. 182; Nagel 1979). The same luck present in the will is likely to lead to the same conclusion about an agent's decisions themselves (Nagel 1979, esp. pp. 26-27). Namely that the agent *did not* exercise sufficient control at the level of decision making. Taken as an argument to show that luck is not a problem, it is in serious danger of backfiring.

On the other hand, we could interpret Kane's argument in the following way. Kane might be asking us to accept that it was not ultimately a matter of luck that the assassin killed the prime minister, because the assassin freely decided to kill the prime minister. Then, we note that the agent freely decided to kill the prime minister despite the chance that he would decide otherwise, just as he freely killed the prime minister despite the chance of failure. But interpreted this way, the argument does nothing to show how ultimate luck has been eradicated, for the notion of what it is for a decision to be freely made is left unanalysed. If Kane means by 'free', made on purpose, intentionally and so forth, then Kane's account of the freeness of a choice is compatibilist, and as such

---

work backwards in exactly the same way, and argue that therefore determinism at the level of decision making isn't a problem

Interestingly a compatibilist could also construct a non-question begging Frankfurt-style case in this way. The compatibilist could ask us to acknowledge that determinism *external* to the agent is not thought necessary by any libertarian version of PAP (the agent does not have a relevant alternative possibility due to the indeterministic possibly of their missing the prime minister due to a jerk in their arm). So, determinism can be safely and non-question beggingly assumed *external* to the agent, thus providing a counterfactual intervener with a reliable basis upon which to intervene. The intervener is not going to intervene at the level of decision making (they've no reliable basis upon which to do so). Rather they will intervene if it becomes apparent that the prime minister will not get shot. If, in the actual sequence, the intervener does not have to intervene, then intuitively this makes no difference to the agent's responsibility for the shooting of the prime minister, even though it was inevitable that the prime minister would get shot due to the intervener's presence. Given the irrelevance of inevitability at the level of events external to the agent's decision making process, and given that such events can impact upon the agent's responsibility, then there is no reason to think that alternative possibilities are needed at the level of decision making either.

I think this version of a Frankfurt-style case (which I do not believe has actually been made, though see Widerker 1995, pp. 253-255, where he seems to suggest the possibility of such a variation, only to dismiss it; and see also Ravizza 1994) is unconvincing – unconvincing for the same reason that Kane's argument is unconvincing.

consistent with ultimate luck. On the other hand, if Kane means by free, non-lucky, then his argument is flagrantly question begging. Clarke makes this point below.

> [T]here is a more fundamental problem in this second appeal to efforts to address the problem of control. In the case of the [assassin] we accept that he acts with the control that suffices for responsibility because we presume that his attempt to [assassinate] is itself free. If, on the contrary, we suppose that the attempt is not free, then we will judge that he does not freely [assassinate]. An effort to make a certain choice can contribute in the same way to that choice's being free, then, only if the effort itself is free. (2002, pp. 372-373, note omitted)

One might object to what has been said so far on the grounds that it is quite *obvious* that on this occasion, it is the second luck objection that Kane is addressing. In other words, Kane is not trying, with the above argument anyway, to show how agent-internal indeterminism *enhances* control, rather he is just trying to show that we should not be so quick to think that it *damages* control, where the control in question is compatibilist. However, whilst I am sympathetic to this point, Kane's subsequent remarks do not support it, as we shall see. For Kane develops the resultant luck argument in such a way as to try and yield a genuine path-picking control. It goes something like this.

First, if we focus on the results of the assassin's act, then what we would say if the assassin missed in the actual sequence, is that he missed by accident, rather than intentionally. The assassin has only one-way control over his shooting of the prime-minister: if he succeeds, he controlled his success, but if he fails, then this was a matter out of his hands. But Kane then goes on to point out that at the level of decision making things are different. We can imagine the assassin deliberating over whether to shoot the prime minister or not – he has reasons for both. It is indeterministic which way he will decide, but the indeterminism is coming from within his will:

> Note that, under these circumstances, the choices either way will not be "inadvertent," "accidental," "capricious," or "merely random," because they will be *willed* by the [assassin] either way, when they are made, and done for

> *reasons* either way [moral convictions if he decides against shooting the prime minister, financial motives if he does] which [he]then and there endorses.
>
> Owing to the indeterminacies in [his] neural pathways , the assassin might miss his target... But [he *succeeds*] despite the probability of failure, [he is] responsible, because [he] will have succeeded in doing what [he was] trying to do. And so it is, I suggest, with self-forming choices, except that in their case, *whichever way the agents choose,* they will have succeeded in doing what they were trying to do because they were simultaneously trying to make both choices, and one is doing to succeed. Their failure to do one thing is not a *mere* failure, but a voluntary succeeding in doing the other. (Kane 2002, pp. 419-421)

So, at the level of decision making the agent has 'two way' or 'stereo' control, or as Kane calls it 'plural voluntary control' (Kane 2002, p. 220). I think that it is fairly clear that Kane is inviting us to see 'plural voluntary control' as path-picking control of the kind that would eradicate causal luck. But having plural control does not involve having *more* control than one would have in a deterministic setting, and certainly nothing that Kane has said above tells us how plural control yields luck *eradicating* control.

Plural control actually amounts to having plural compatibilist control, meaning that whichever way the agent decides on the occasion of a torn decision, the resulting decision will have been made for reasons, purposefully etc. Yet Kane has a tendency to talk as if plural control is a kind of luck eradicating control, or at least he seems to come perilously close to saying this kind of thing. For he says things like plural voluntary control involves having "the ability to cause or produce any one of a set of possible choices or actions each of which is undetermined ... and to do so "at will" (that is, rationally (for reasons), voluntarily, and intentionally)" (2002, p. 431). It is the word 'ability' that is troublesome here. For to use it is to imply that the agent can antecedently determine which choice they make. In other words, it looks to me as if Kane is trying to slip-in a special form of agency or causation of precisely the kind associated with the extra factor incompatibilist views he rejects.[74] Kane expressly rules this out, but to at the

[74] Hodgson also thinks that this is what Kane does – though Hodgson *endorses* this kind of amendment to Kane's view, whilst recognising that Kane himself would want to reject the charge (Hodgson 2002, pp.

same time talk about undetermined choices as delivering an extra 'ability' to the agent that they would lack in a deterministic setting is at best extremely misleading, and at worst downright false. Yet Kane does persist in talking about plural voluntary control as if it offered something more in the way of control than compatibilists offer (2002, p. 431):

> I think the "more" control libertarians need is not more of the same *kind* of control compatibilists offer, but rather another kind of control altogether... What libertarians must require for undetermined SFAs [Self-forming acts], I believe, is another kind of control altogether (that compatibilists cannot obtain) – namely, *ultimate* control – the originate control exercised by gents when it is "up to them" which of a set of possible choices or actions will now occur, and up to no one and nothing else over which the agents themselves do not also have control... It [involves] the ability to cause or produce any one of a set of possible choices or actions each of which is undetermined (hence non-deterministically) – and to do so "at will" (that is, rationally (for reasons), voluntarily, and intentionally. (Kane 2002, p. 431)

But Kane's use of the word control here is misleading in a way already noted when discussing Valerian incompatibilist positions. Plural voluntary control involves having, on occasion, exclusive control in the sense that nothing external determines which way one will decide, yet whichever way one decides one will have decided for reasons, on purpose and so forth. But this exclusivity does not yield an increase in the agent's control in any absolute way, and certainly not in any way that could rule out ultimate luck. For as with the Valerian view given earlier, the kind of 'exclusive' control that the agent has can only be compatibilist control. Nothing has been said which shows how the agent could have anything beyond this.[75]

109-110 and see also his 1999, p. 214. Kane discusses Hodgson's view in his 2002, p. 434 n. 15, where he confirms that he does not take himself to be introducing any kind of extra factor or non-event causation).

[75] Balaguar is another event-causal incompatibilist in Kanean mould. When considering the 'luck' objection, Balaguar replies as follows:

> No, it's not *just* a matter of luck or chance. It's true that, in some sense, Ralph's decision was arbitrary (or if you like, chancy or lucky). But this chancy event was still a *decision* (it was conscious, intentional, purposeful, and so on) and it was still *Ralph's* decision: nothing external to him made him choose as he did. So it's not obvious that Ralph didn't author and control the decision. This is a delicate matter. The question is whether the intuitive notion of *being in control of what you do* applies in this case. I want to argue that if Ralph's decision was

It is, I hold, very important to distinguish between the claim to have given an agent exclusive compatibilist control, and the claim to have supplied the agent with more than just compatibilist control. Blurring this distinction leads to nothing but confusion. It leads modest incompatibilist such as Kane and others, to think, or suggest, that they've got something like a solution to the problem of ultimate luck. It makes them think that they've got a way of showing how all that we do is not a matter of luck, without having to hide behind mysterious forms of agent-causation. But as I hope my discussion above has brought out, such accounts, whether of the Valerian or Kanean modest variety, do no such thing.

5.13 Conclusion

In this chapter I have argued that if incompatibilist path-picking control is understood in terms of compatibilist antecedent ensuring control then it is incoherent. If it is not understood in this way, then it is ineliminably mysterious. Perhaps some are content to appeal to mystery, but this will not persuade those not already committed to incompatibilism.

Modest 'event causal' incompatibilists make no recourse to mysterious extra factors, and instead present accounts of free will in terms of normal physical and psychological processes. But their modesty costs them any right to have provided anything in terms of control over what one would have in deterministic settings. As such, where ultimate luck is concerned, one is no better off under indeterminism.

I believe that those who defend such accounts – Kane, Balaguar, Ekstrom, Nozick, and others – are going to be, in general, reluctant to admit this, or will do this in a rather coy and ambiguous way. But I think this reflects the fact that (and this is a guess) the above

---

> undetermined at the moment of choice, then the intuitive notion of control clearly does apply (and that the notion of authorship does as well), despite the fact that the choice was, in some sense, arbitrary or chancy or lucky. (2004, p. 389)

Balaguar is not clear, but from the above quote it does seem to me that Balaguar is conceding that in terms of control one has no greater control than in deterministic settings. For if determinism were true all of Balaguar's above claims would also hold true.
So long as that is the case, what the agent does is as much a matter of luck as it would be in a deterministic setting.

have probably arrived at their modest incompatibilist positions from former control incompatibilist positions. As such they retain a deep seated conviction that compatibilist positions are inadequate partly because of the quality of the control they offer. For such reasons these modest incompatibilists are still eager to show how their accounts offer something *extra* in terms of control.[76]

But, I hold that we make greater progress by being honest, and owning-up that ultimately everything we do is ultimately a matter of luck irrespective of whether agent-internal indeterminism obtains or not. It is a revelation that modest incompatibilism can survive, and can survive in better shape than compatibilism. Or so I will argue in the coming chapters.

[76] Nothing prevents proponents of modest incompatibilist views appealing to mystery in the same way that 'extra factor' incompatibilists do. After all, strictly speaking it is no less mysterious how an extra factor, such as a special form of causation, could yield path-picking control anymore than Kane's plural voluntary control could yield such control. If one is going to rule out causal luck by fiat, one can do it just as well in the context of an event-causal view as one can an agent-causal view.

Chapter 6

Indeterminism and Control: II

The conclusion of the previous chapter was that we are no better off in terms of control if our decision making processes are indeterministic than we would be if they were deterministic. But might indeterminism actually make matters worse in terms of control? In other words, is the 'erosion' version of the luck objection correct?

In what follows I will argue that agent-internal indeterminism does not damage control, or at least, it does not damage control in any responsibility subverting way. In terms of control we may be no better off with indeterminism, but we are no worse off either.[77]

Making the case against the erosion argument will concern me for the bulk of the chapter. However, concerns about indeterminism damaging control are not the only respects in which indeterminism is seen as a danger to our status as free, responsible agents. I will consider three further concerns which have to do with 'explanation', 'rationality', and 'attributability'. I will argue that the concerns here are misguided.

I will start with a couple of preliminary remarks. If we understand the erosion argument to be the claim that in terms of control matters would be better if determinism were true, then the kind of control indeterminism is eroding must be compatibilist control. By definition, that is the only kind of control that one can have in deterministic settings. Relatedly, when I talk about control not being eroded in a way that would subvert responsibility, I assume that I am addressing a compatibilist. By hypothesis, a control incompatibilist thinks we need *more* than compatibilists control for responsibility and so would think responsibility subverted irrespective of how much compatibilist control we have in deterministic settings.

---

[77] The 'erosion' version of the luck objection has been made in one form or another by: Bernstein 1995; Double 1991, pp. 198-199; Fischer 1999b; Haji 1999a & 2002; G. Strawson 2003; Waller 1988. I focus on Haji's version because Haji does at least couch it in terms of luck.

6.1 Valerian Incompatibilism

In the previous chapter I distinguished Valerian incompatibilism from other kinds of modest incompatibilism. Although I am not a Valerian incompatibilist, I think that it is clear that such views locate indeterminism in a place where it does not erode an agent's compatibilist control.[78] For the Valerian incompatibilist, to remind ourselves, locates indeterminism in the coming to mind of beliefs, and we do not have control over such matters even if determinism is true (and nor would we if we had incompatibilist control). An agent can hardly be described as worse off in terms of control if the indeterminism is located in a place over which we never would have had any conscious control of any responsibility relevant kind under determinism. However, not all agree. I will consider a couple of objections.

Clarke has recently pointed out that the introduction of indeterminism in this way might be said to erode or damage a certain type of "nonactive" control. He explains:

> It could be that, when an agent sets out to make up his mind about which of several alternatives to pursue, all and only the most relevant considerations, or all and only those that he has time to consider, come promptly to mind and then figure rationally and efficiently in the production of an evaluative judgement. In a deterministic world in which our deliberations always ran in this ideal fashion, we would exercise a valuable type of nonactive rational control in deliberating. (2002, p. 380, n. 19)

I actually think that this is correct and I return to this kind of concern later in the chapter. But at the moment it suffices to say this: the issue is whether indeterminism would damage an agent's control in a way that could plausibly subvert responsibility. It

---

[78] The Valerian could claim that indeterminism early in the actional pathway is one amongst a set of jointly sufficient conditions for responsibility grounding free will. As such it is quite consistent for a Valerian incompatibilist to allow that indeterminism elsewhere is not a problem. A Valerian could allow that if indeterminism obtained in the way outlined by Kane's account, and *not* in the 'coming to mind' of beliefs, that this too is sufficient for responsibility. I do not think that any actually say this – for the main motivation for subscribing to Valerian incompatibilism is precisely the concern about control-erosion voiced about views such as Kane's. But the point is that they could say this, for their motivation might primarily be a concern about securing ultimacy. And it is plausible that one could see both Valerian and Kanean modest incompatibilist views as equally good on that front. This is something I actually take issue with in Chapter 7 where I argue that Valerian views *cannot* secure ultimacy or 'sourcehood' in a way that would deliver the kind of ownership over what we do that we saw determinism subvert in Chapter 4.

may be the case that indeterminism, even located in what otherwise appears to be the entirely safe place that the Valerian incompatibilist locates it, precludes possession of the type of ideal, nonactive rational control Clarke outlines above. But no plausible position on free will can hold that ideal nonactive control is required for responsibility-grounding free will. After all, it certainly does not follow from determinism being true that we have ideal nonactive rational control, and nor would it follow if, per impossible, we could possess incompatibilist control. In both cases our exercise of responsibility-relevant control comes after beliefs have come to mind – we exercise control over how we respond to such matters, rather than over such matters directly. If we fall short in terms of ideal nonactive rational control – as we surely nearly all do – then it is hard to see how indeterminism is doing any responsibility-subverting damage.

Fischer has also charged that Mele's Valerian view erodes control:

> [E]ven though the agent does not directly control what belief-states come to mind (in the sense of choosing them or willing them), they are envisaged as strongly connected to the agent's prior states to the extent that they are a *deterministic product* of those past states. Under determinism, one's prior states – desires, beliefs, values, general dispositions – *determine* the precise content and ordering of the subsequent doxastic states (that constitute deliberation), even if the agent does not directly *control* what doxastic states he will be in... It may then be possible to argue that one *does* give up some measure of control, when one shifts from thinking of the doxastic sequence as deterministic to thinking of it as indeterministic: one gives up the notion that the states constituting one's deliberations re an "outflowing" of the agent's prior states in a strong sense. (1999b, p. 141)

However, we can just reiterate that on any compatibilists view the agent exercises their responsibility relevant control after their belief-states come to mind. It is true that intuitively, an agent can plausibly be considered responsible for what beliefs came to mind during some episode of deliberation, if, that is, they had earlier made themselves such that certain beliefs did not come to mind. For instance, I often infuriate my partner whenever she blames me for something I have failed to do by pointing out that it "just didn't occur to me" and that I cannot be blamed for what did not occur to me. This

excuse rarely works, because she charges that there was some earlier time when I did not take steps to make myself such that these kinds of thing would occur to me. So, my responsibility for what beliefs occur to me traces to earlier free acts. If I had made efforts to make myself such that certain belief states do occur to me, and indeterminism is then introduced into the processes that bring beliefs to mind, then we could see such indeterminism as eroding my control. My good work is undone. The indeterminism serves as an obstacle to my being able to antecedently ensure that certain beliefs occur to me. But this kind of erosion of control cannot be seen as subverting responsibility, for that my responsibility ultimately traces (on a compatibilist view) to my directly free acts – acts which do not involve any exercise of control over what belief-states come to mind.[79]

We can make the point another way. Imagine that it is deterministic what belief-states come to mind during my deliberations but I do try to make myself such that certain beliefs will occur to me on certain occasions. Unfortunately, all my attempts are doomed to failure – it is deterministic that certain beliefs will never occur to me on the relevant occasions, irrespective of any prior efforts I may have made.[80] Well, unfortunate as that may be, it cannot plausibly destroy my responsibility for how I respond to what does occur to me on those occasions.

So Fischer's criticism does not highlight a way in which indeterminism would erode control in any responsibility subversive way. It appears quite clear that the truth of determinism would not guarantee that one had the type of control in question. In short, if indeterminism is located in the coming-to mind of certain beliefs, then in terms of responsibility-relevant control, we are no worse off than under determinism.

6.2 Non-Valerian Modest Incompatibilism

As I pointed out in the previous chapter, the luck objection has, in recent years, been directed for the most part at Kane's brand of modest Incompatibilism. When it comes to

---

[79] Haji makes essentially the same point in his 2002, pp. 98-100. Or at least I think he is – I found his discussion on this point a little hard to follow, but it did prompt the points I am making here, irrespective of whether they accurately reflect what Haji actually meant.

[80] Note, I am not committing the fallacy of thinking that determinism *entails* that my efforts to make myself a certain way are doomed to failure. I am just pointing out that it could be the case that, as it

the erosion argument is has to be admitted that matters are slightly murkier than with the Valerian view. However, I think that a good case can be made for saying that indeterminism located between the agent's reasons and decision does not damage an agent's control in any responsibility-subverting way, so long as attention is paid to comparing relevant cases.

The compatibilist Ishtiyaque Haji, is prominent amongst the large number who disagree (1999a; 2002, sec. 2). Haji maintains that where indeterminism is just prior to decision making, then what the agent lacks is antecedent proximal control, which is control "*to see to it that, in that rerun, he smokes rather than that he does not* and vice versa" (Haji 2002, p. 110).

> With *fixed pasts*, the difference in outcome in Jones's and Jones*'s cases [these are the Jones and Jones* of the reruns in Haji's version of the luck objection] appears to be merely a function of the indeterminacy in the actional pathways leading to choice. But it would seem that no agent could exert proximal (or any other sort of) control over such indeterminacy to ensure a particular outcome. (Haji 2002, pp. 110-111)

According to Haji, the relevant, destructive difference between an agent's torn decision being determined to be resolved in a certain way, and it being indeterministic which way it will be resolved, is that in the latter case the agent cannot antecedently ensure the outcome. This is an ability that Haji thinks the agent would have in a deterministic setting.

There are a number of responses we can make. First, adopt for the sake of argument a straightforward compatibilist position of the type that Haji and many other contemporary compatibilists endorse. Genuine alternative possibilities are not needed, and it is enough that in the actual sequence the agent exercised compatibilist control over the decision that they made, where this is a matter of the right connections being (the agent decided for reasons and their response was an output of a sufficiently reason-responsive mechanism operating against a suitable background evaluative scheme).

---

happens, my efforts will fail because this aspect of my mental economy is impervious to my attempts to alter it.

Now introduce agent external indeterminism into the picture in the following way. Due to certain properties of the agent's brain it is indeterministic whether or not the agent will suffer a massive stroke at t1. As such there is now a genuine possibility that, rather than make the decision in question at t1, the agent will die of a stroke at t1. In one respect this clearly diminishes the agent's antecedent determining control. The agent cannot, prior to t1, ensure that they make a particular decision at t1, for they cannot ensure that they do not suffer the stroke. If we replay the universe innumerable times, sometimes the agent will make a decision at t1, sometimes they will be dead at t1.

Yet it seems quite obvious that if, in the actual sequence, the agent makes the decision in question, then they are as morally responsible as if there had been no possibility of anything else happening. After all, in the actual sequence in which the agent makes the decision, all the right connections were made and so the compatibilist control requirements are met. Whilst the external indeterminism could be said to serve as an obstacle to the agent successfully making a decision at t1, it does not serve as a responsibility-subverting obstacle.

Now we can note that if the indeterminism is internal to the agent's decision making processes rather than external, that does not make any difference in terms of the reliability with which the agent can ensure a particular decision is made. In other words, if it is indeterministic whether Jones will make decision X or be dead at t1, or indeterministic whether Jones will make decision X, or decision Y, Jones's ability to reliably ensure that he makes decision X is equally damaged in both cases. Jones can no more antecedently ensure that he makes decision X at t1 when the source of the control-diminishing indeterminism is *external* to his will, than he can when it is coming from within his own will.[81]

---

[81] Kane can also be seen to be making this kind of point:

> Suppose you are trying to think through a difficult problem, say a mathematical problem, and there is some indeterminacy in your neural processes complicating the task – a kind of chaotic background... Whether you are going to succeed in solving the mathematical problem is uncertain and undetermined because of the distracting indeterministic neural noise. Yet, if you concentrate and solve the problem nevertheless, we have reason to say you did it, and are responsible for it even though it was undetermined whether you would succeed. (2002, pp. 417-418; see also 2003, p. 308)

However, as pointed out in the previous chapter, Kane tries to use this kind of point to make a case for indeterminism enhancing control, when in fact what he has done is highlight how it does not erode control.

My point here parallels a point I made in the previous chapter. I argued that moving indeterminism inside the 'will' cannot plausibly be said to enhance an agent's control any more than external indeterminism can. Here I am saying that similarly, indeterminism cannot be said to diminish an agent's control when it is moved inside the will anymore than it can be said to diminish the agent's control when it is external to the will. As I have already acknowledged above, there is a case for saying that indeterminism diminishes control irrespective of whether it is internal or external to the will, precisely because one can less reliably ensure certain outcomes. But my point is that unless one wants to be committed to the claim that external indeterminism rules out responsibility, one cannot claim that internal indeterminism is fatal to responsibility-grounding compatibilist control.

One might object here, that my above argument is relevantly identical to Kane's resultant luck argument given in the previous chapter, and as such I am vulnerable to the same criticisms that I myself made. For surely, one might say, external indeterminism *is* a problem insofar as it makes it chancy whether we will succeed in our actions, and so what my above argument depends upon is prior acceptance of the compatibility of luck and responsibility. In response, I agree – that is what I am asking. But unlike before it is legitimate for me to do so here, for I take it that those running the erosion argument will be compatibilists, and as such they will have to accept the compatibility of luck and responsibility. After all, compatibilist control does not provide protection against ultimate luck, irrespective of whether determinism or indeterminism is true.

A different response might be to point out that there is a relevant difference when the indeterminism is located between the agent's reasons and decision. For if the indeterminism is external, then although the agent cannot strictly speaking ensure that they make (say) decision X at t1 (for they might be dead at t1), they nevertheless can ensure that if they make a decision at t1, it will be decision X.

But there is something odd about saying that the agent can ensure that they make one decision rather than another, when the control in question is compatibilist. For that implies that we decide to decide – that we choose to deploy some kind of decision-making ability. This *isn't* the case at the level of decision making. Or at least, we do not

ultimately do this. That is the kind of thing we do through decision making, not prior to it. Randolph Clarke has made this point:

> [W]e need to distinguish the following two significantly different varieties of case: those in which there is indeterminism between a basic action and an intended result that is not itself an action, and those... in which the indeterminism is in the production of a basic action itself. For the first sort of case, suppose that you throw a ball attempting to hit a target, which you succeed in doing. The ball's striking the target is not itself an action, and you exercise control over this event only by way of your prior action of throwing the ball. Now suppose that, due to certain properties of the ball and the wind, the process between your releasing the ball and its striking the target is indeterministic. Indeterminism located here inhibits your success at bringing about a nonactive result that you were (freely, we may suppose) trying to bring about, and for this reason it clearly does diminish your control over the result – it constitutes control-diminishing luck. But the indeterminism... required by the sort of event-causal libertarian view at issue here [Kanean modest incompatibilism] - is located differently. It is located not between an action and some intended result that is not itself an action, but rather in the direct causation of the decision, which *is* itself an action. The control that an agent exercises in making a decision does not (typically) derive at all from any prior attempt on her part to bring about that decision. In the ball-throwing case, the indeterminism constitutes control-diminishing luck because it inhibits the agent from bringing about a nonactive result that she is actively trying to bring about. But that explanation is not available in the second kind of case. Unless the argument from luck offers some alternative explanation, that argument is at best inconclusive. (2002, pp. 367-368, two notes removed)

As Clarke points out, external indeterminism seems to diminish control because of its location between the agent's decision and their action, whereas this is not where the indeterminism is located on Kane's view. The disanalogy means that there is a genuine question mark over whether internal indeterminism can correctly be said to diminish control.

6.3 Indeterminism and Reason-Responsiveness

I think we can go beyond what Clarke as said above. Note that the agent-internal indeterminism, located between reasons and decision, is consistent with the agent's decision having been the output of a reason-responsive mechanism, and so the compatibilist needs to give us some reason why exactly compatibilist control conditions are not met.[82] On a plausible account of reason-responsiveness, we ask whether holding fixed the relative strength of the actions motivational precursors and the agent's background evaluative scheme, there are situations in which a different decision would be issued. This condition is satisfied in the case of an indeterministically resolved torn decision, for the agent will sometimes make a different decision in roll-back cases – cases in which all prior conditions are held fixed, including motivational precursors and evaluative schemes.

What we need to remember is that on the model of compatibilist control (to which Haji subscribes) what matters is the quality of the connections made in the actual sequence, rather than what could have happened in the actual sequence. So, unless it is simply stipulated that one of the quality requirements is that an agent's decision must be deterministically caused by their prior reasons there is no reason to think that agent-internal indeterminism jeopardises this kind of compatibilist control. If Haji does say that a deterministic causal relation is what is required, then he has just begged the question against the Kanean incompatibilist.

---

[82] I focus on the reason-responsive account of compatibilists control. But it is even more straightforward that on other compatibilist accounts control conditions are met in the relevant cases. For example, as we saw in Chapter 4, hierarchical 'mesh' accounts hold that an agent's act of will was appropriately controlled just if the right 'snap shot' properties obtained, where this is a matter of higher and lower order desires 'meshing'. Well, it does not matter how this mesh came about – that will not affect the snap-shot properties.
One might object that hierarchical views are 'harmony' views, where 'meshing' involves first order and second order desires agreeing in some appropriate sense. This, it might be objected, is not what one gets in a torn-decision, for by definition the agent's will is divided on such occasions. But to this we can make two replies. Firstly, a hierarchical account can hold that an agent controls their decision when it accords with their values, and if one of the possible decisions which they are 'torn' between would, if made, accord with their values (this would be the decision which the agent judged to be the best one on the basis of her valuational system, though whether she will actually make this decision will depend upon her *motivational* system (see Watson 1975)) then *that* one would be controlled if made. In other words, one would have here a 'one-direction' control (the decision would be controlled if it turned out one way, but not if it turned out the other). Secondly, whatever problems torn-decisions might turn up for a hierarchical view (or any other compatibilists view), these problems stem from the 'torn' nature of the decisions rather than whether or not they are deterministically or indeterministically resolved.

The point can be made in a different way. The erosion argument gets a lot of its superficial credibility from not comparing relevantly similar cases. What needs to be borne in mind is that on the Kanean view it is only our torn decisions that are indeterministically resolved, for it is only on the occasion of a torn decision that indeterminism is stirred-up in the will. Assuming that there is absolutely no reason to think that we do not face torn decisions if determinism is true, then the cases we need to compare are torn decisions that are resolved indeterministically and torn decisions that are resolved deterministically, and where this is the only difference between them.

Take a very crude model of compatibilists control (my argument would apply equally to more sophisticated versions). On this model an agent's decision is appropriately controlled if he would have decided otherwise had he wanted to. To stress, it does not matter whether the agent actually could have wanted to decide otherwise in the actual circumstances in which they make the decision, for all we are doing is assessing the responsiveness of the mechanism leading to the agent's decision. But note that in the case of a torn decision we encounter problems. For in the case of a torn decision, the agent wants to make two incompatible decisions, but can only actually make one. This model of control cannot say that the agent would have decided otherwise if she had wanted to, because by hypothesis she did want to decide otherwise *in the actual sequence*. Left un-amended this account of control would therefore deliver the curious result that an agent does not control their torn decisions, even in deterministic settings.

So, unless the compatibilists wants to be in the embarrassing situation of having to say, counter-intuitively, that we are not responsible for our torn decisions, (irrespective of whether they are resolved deterministically or indeterministically) then the compatibilist is going to have to endorse a notion of 'moderate' responsiveness generous enough to include torn decisions.[83] In other words, the compatibilist will have to say that an agent's decision issues from a suitably responsive mechanism provided there are a sufficiently broad range of circumstances in which the agent would have made a different decision if there had been reason to, without committing themselves to the far stronger claim that the agent would always decide otherwise if there was reason to. But now note that if the compatibilist makes their reason-responsive conditions moderate

[83] As pointed out in Chapter 2, any plausible reason-responsive account of compatibilists control is going to have to be 'moderate' if it is to yield the result that we are ever blameworthy.

enough to cover torn decisions, then indeterministically resolved torn decisions will satisfy such control conditions too.[84] There is no way for the compatibilist to avoid this without just stipulating that a process or mechanism leading to decision needs to be deterministic. In other words, there does not seem to be any non question begging way in which the compatibilists can show why indeterministically resolved torn-decisions fail to satisfy moderate reason-responsiveness requirements.

There is another way of making the case for will-internal indeterminism not undermining control, different from that above, but which again depends – quite rightly – upon comparing deterministically resolved torn decisions and indeterministically resolved torn decisions. In the previous chapter I used an example of a TV remote to illustrate indeterminacy in the agent's effort of will. I now want to return to that example. Imagine you are offered a choice of two TV remotes for your television. Both are equally reliable if you press one button at a time. But if you press two buttons at the same time on the 'Kane' remote control, then it will be indeterministic to which channel the remote will switch (though it will switch to one of the two selected). Whereas if you do the same on the 'Haji' remote control then although you will not know which of the two channels it will change to, it will be deterministic which channel it will change to.

One might object that if the remote control operates deterministically, then we can know to which channel it will change if two buttons are simultaneously pressed, provided we have enough information about the environment and internal workings of the mechanism on that specific occasion. This may be true in principle, but I rule it out here because when we are subject to a torn decision we do not know which way we are going to resolve our decision until we actually make the decision in question, irrespective of whether our torn decisions are deterministically resolved, or indeterministically resolved. I take it that this is a fact about torn decisions – if one knew how one was going to resolve one's torn decision one would thereby have resolved it (Bok 2003, p. 157; Perry 1965, p. 239). The stipulation that we do not know which way the Haji

[84] The compatibilist might object that in cases of torn-decision making there is not *one* mechanism or process operating, but two. This interpretation is lent support by Kane's own conceptualisation of torn decisions as involving parallel processing. But this will not help the compatibilists, for if we just focus on the mechanism that issues in the decision in the actual sequence, then this mechanism will report the same degree of responsiveness irrespective of whether, in the actual sequence, the *dual* process was one that contained indeterminacy. For the *other* process is now to be thought of as an *external* source of interference. And external indeterminism does not affect reason responsiveness.

remote will change on every occasion in which two buttons are simultaneously pressed, is warranted.

Given this, which remote would you pick? Which remote would you judge to give you greater control over your future channel-changing needs? It seems to me that there is no relevant difference between them. Both work equally well when just one button is pressed at a time, and when two buttons are pressed at once, one is in no better position to know which way the 'Haji' will change channels than one is with the 'Kane' channel changer.

Imagine that God gives you a choice between two different kinds of decision-making control mechanisms. Whichever one you choose, it will be the mechanism by which you make your future decisions. There is absolutely no difference between the mechanisms, except when it comes to torn decisions. As with the TV remotes, with the 'Kane' mechanism installed it will be indeterministic which decision you will make on the occasion of a torn decision, whereas with the 'Haji' mechanism it will be deterministic. But in both cases you do not *know* which decision you will make until such time as you actually make it. And nor do you know prior to having the mechanism installed, which way you will decide when it comes to torn decisions (but we can assume that you will face some torn decisions and that you know that you will). So which kind of control mechanism would it be wisest to have installed, assuming that is, that one wants to maximise one's control over one's future decisions? Again, it seems to me that there is no relevant difference between the two. By my lights the 'Haji' does not give one any greater control than the 'Kane'.[85]

---

[85] Balaguar has recently argued something similar. Balaguar has argued that if one focuses on torn decisions in deterministic settings, then what explains why the agent decided one way rather than another must be something "extra-agential":

> [S]ince the agent's reasons don't pick out a unique best option, anything that dose determine this (e.g., a non-mental brain event in the agent's head) would presumably be extra-agential, and so if our torn decisions are determined then they are not under our control. (Balaguar 2004, p. 391)

But Balaguar wants to claim something stronger than I do. Balaguar thinks that indeterminism in a torn decision is "the only way to guarantee control for the agent" (2004, p. 391). I certainly don't think that is right – though my disagreement with Balaguar on this point may be more terminological than anything else (and to some extent this may be true with respect to Kane too). For Balaguar uses the word 'control' in such a way as to include what I would want to distinguish as utlimacy.

If this is right it seems to me that there is no reason to see indeterministic efforts of will as diminishing of control in any way that matters. One might want to say that torn decisions are less controlled than non-torn decisions, precisely because when we are torn we do not know which way we are going to decide until we actually *do* decide. But this would be a point about torn decisions, irrespective of whether they are resolved deterministically or indeterministically.

6.4 Kane and Balaguar's Intuitive Argument

We can add to what I have said above the following intuitive argument, versions of which have been made by both Kane and Balaguar (Kane 1996, pp. 182-183; Balaguar 2004, p. 393). Imagine you are subject to a torn decision, and eventually you just decide one way. Your brain was being monitored whilst you were making this decision. The doctor monitoring your brain tells you that, according to their readouts, it was literally indeterministic which way you would resolve your torn decision, up until the moment you actually made the decision. Upon hearing this would you conclude that you didn't control your decision? I wouldn't. And Balaguar thinks it "would be downright bizarre" to think that one didn't control it (2004, p. 393).

6.5 Retentive Control

So far I have argued that we have no reason to think that we will have less responsibility-relevant control over our torn decisions if they are resolved indeterministically than if they are resolved deterministically. However, I now want to turn to one respect in which we might feel that our control is diminished. Earlier, both Clarke and Fischer pointed out that there may be a case for saying that indeterminism introduced in the way typical of Valerian incompatibilism, might diminish certain varieties of non-active control.

I think that a similar concern can be voiced about Kanean modest Incompatibilism. Consider that one might like the way that one currently is – one might identify with one's values and so forth – and one might want to remain as one is. The prospect of future SFAs where it will be indeterministic in which direction one's future development takes, will therefore be of concern. For these future SFAs on the horizon

threaten to change you. These future SFAs seem to pose a threat to what I term 'retentive' control. I want to know that I will always act in certain ways in certain situations, but SFAs deprive me of any guarantees.

I think that this is a legitimate worry. There are some, perhaps many aspects of our characters that we want to ensure do not change, and if SFAs are a necessary part of what it is to be free, then free will entails that our characters are *unfixed* to some extent, and liable to alteration. But we should note two things. Firstly, whilst the prospect of having an unfixed nature can be a cause for distress, so too can the opposite. The idea that our nature is fixed is one that we are likely to find just as unpalatable. Often determinism is (wrongly) identified with having a fixed nature, and for that reason is seen as unpalatable. This is a point that I am going to be returning to in the next chapter.

The second point is that whilst determinism gives one the prospect of having a fixed nature, it certainly does not guarantee it (it is this mistake which people make when they wrongly identify determinism with having a fixed nature – as Dennett points out, determinism is the thesis that one has a fixed future, not a fixed nature). There is no reason to think that if determinism is true, that one's virtuous nature will remain fixed. There is no reason to think that just because determinism is true we will not face torn decisions where we are conflicted about the kinds of people we want to be. Again, one seems no worse off if one's torn decisions are indeterministically resolved than if they are deterministically resolved. In terms of retentive control, the risk is posed by torn decisions, whether of the deterministic or indeterministic kind.

That completes my discussion of the erosion argument as it pertains to control. There are types of non-active control that both Valerian and Kanean modest incompatibilism can be seen to threaten, however such types of control can be equally threatened in deterministic settings, and more importantly, these types of control are not required (above some threshold level) for responsibility. With respect to responsibility-relevant compatibilists control – control in how one reacts to what comes to mind, and control in how one resolves one's torn decisions – there is no reason to see the indeterminism as depriving one of any relevant degree of control.

However, as I mentioned above, the concern about indeterminism doing damage is not limited to concerns about control. I now want to consider three other concerns about the availability of relevant types of explanation; concerns about the rationality of the agent's decision, and one concerns attributability.

6.6 Contrastive Explanation, Rationality and Attribution

Haji, a compatibilist, is prominent amongst those who make the 'erosion' argument against Kanean brands of incompatibilism. But Haji thinks that the problem is "intimately tied to lack of an explanation in terms of prior reasons of the difference in choices" that an agent makes in "re-runs" of torn decisions. He is not alone. Many others have couched the erosion concern in terms of the lack of a certain type of explanation of why the agent decided as they did, rather than otherwise (Ayer 1954 pp. 3 - 20; Double 1996; and Nagel 1986 pp. 113-117). The explanation in question is a 'contrastive explanation', which is to say an explanation of why one thing, rather than another, occurred.

However, contrastive explanations are not the only kind of explanation that there is. There are also 'plain' explanations. A plain explanation is an explanation in terms of the events prior causes. If an event is indeterministically caused, it is still caused, and as such a plain explanation is still available.

Returning to the internally indeterministic remote control – the 'Kane' remote - even when both BBC1 and BBC2 buttons have been simultaneously depressed we can give a plain explanation of why the remote changed channel to, say, BBC1, by citing the fact that the BBC1 button was pressed. We cannot explain why the remote changed channel to BBC1 rather than BBC2, for prior conditions were consistent with either of these events occurring. So a contrastive explanation is ruled out (note, not all contrastive explanations are ruled out – we can still contrastively explain why the remote changed channels to BBC1 rather than ITV for example).[86] This will be so in the case of SFAs too. So Haji is correct – we cannot explain why Jones decides one way in some re-runs,

[86] Clarke has argued that we can still provide adequate contrastive explanations in indeterministic settings. See his 1996.

and a different way in others.[87] But the important point is that the lack of contrastive explanation does not rule out the availability of a plain explanation.

Why should the lack of a relevant contrastive explanation be a problem in itself? It would not do just to insist on it. For if contrastive explanations are not available in any setting other than one in which an agent's decision is determined by their prior reasons, then this just begs the question against the modest incompatibilist.[88] It must rather be that the lack of contrastive explanation is taken to indicate either a lack of control – in which case the arguments I have made in the preceding sections take over - or it might be taken to indicate a lack of rationality or a problem with respect to 'attribution'. (The latter is, it seems, how Haji intends this point, as we shall shortly see.)

There is a concern that with the indeterminism located between the agent's reasons and their decision, the decision made is insufficiently connected to the agent's reasons for action (see Double 1996, pp. 69-76 and see also his 1991)

However, we've already noted above that for the agent's decision to have been indeterministic does not mean that the decision that was made will not have been antecedently caused. If an agent's decision making process is indeterministic up to the moment of decision, that does not mean that we cannot give a plain explanation of the agent's decision in terms of the reasons that caused it in the actual sequence. So there is no reason to think that we will not be able to link the agents' decision with their prior reasons.

Another way to put this is to say that whichever way the agent decides their decision will have *teleological intelligibility*, at least in the relevant cases. Following Kane what we can say in cases where the agent has competing reasons for action, and it is indeterministic which decision the agent will make, then it is nevertheless the case that whatever decision is made it is made for reasons, and so can be described as 'dual rational':

---

[87] This is something Kane acknowledges, as does Sorabji – another event-causal incompatibilist (Kane 1996, p. 145; Sorabji 1980, p. 31; see also Fischer 1999a, p. 103 n. 18)

[88] Another equally question begging criticism, which I've chosen not to make part of the main text of this chapter, concerns a claim about choices themselves. For it might be argued by some that 'choices' cannot be identified with undetermined events (see for instance, Bernstein 1995, p. 154). Undetermined events

> The choice of A y an agent is *dual rational*, if and only if, *whichever way it goes*... the outcome is a) the intentional termination of an effort of will that is the agent's... b) the agent has reasons for the choice (whichever occurs), c) the agent does it *for* those reasons, and d) given the agent's character and motives, it is, all things considered, rational for the agent to do it at that time for those reasons. (Kane 1988, p. 446)

Nevertheless, the critic might insist that unless the agent's decision is entailed by their reasons, then their decision is not rational. Following Balaguar we can call such decisions *strongly* rational (2004, p. 395). This is as opposed to a *weakly* rational decision.

> [A] decision is *strongly rational* (given the agent's reasons for choosing) if and only if the agent's (conscious) reasons for choosing entail that the option chosen is the best available option. Second, a decision is *weakly rational* (given the agent's reasons for choosing) if and only if it is consistent with the agent's (conscious) reasons for choosing. (notice that weak rationality is certainly a kind of rationality; in particular, a decision that's weakly rational is not *irrational* in the sense of going against the agent's (conscious) reasons for choosing.) (Balaguar 2004, p. 395)

In the case of a torn decision the decision will clearly fail to be strongly rational, because the decision will not have been uniquely picked out by the agent's reasons. However, the decision will be weakly rational, for it is going to be consistent with the agent's reason set. The question then becomes whether weak or strong rationality is required for responsibility. For if only weak rationality is required for responsibility, then the indeterminism central to the robust modest libertarian view does not threaten responsibility. But as Balaguar has pointed out, strong rationality is clearly not a requirement of responsibility:

> Suppose someone said the following... Jane did not have compelling reasons

---

are just things that happen. As Kane points out though, this implies that "if anything is a choice or action, it is determined" (2002, p. 423). This just rules out incompatibilism by fiat.

> that uniquely picked out her choice of tiramisu, so she did not choose the tiramisu of her own free will, because her choice was not sufficiently rational... it's pretty clear that this remark would seem downright bizarre to just about everyone. In our ordinary discourse and thought, we simply don't think that because decisions like Jane's aren't strongly rational, they are not free. (2004, p. 396)

So, whilst agent-internal indeterminism does preclude strong rationality for torn decisions, strong rationality is not plausibly a requirement of responsibility. If weak rationality is required (and it is not clear it is), then the relevant torn decisions satisfy the weak rationality requirement.[89]

6.7 Attributability

A different kind of concern is that the agent-internal indeterminism means that we cannot connect what the agent actually decides with any fixed character. When an agent's will is indeterminate, there does not seem to be any stable character to attribute their actions to. Consider Haji's comments below:

> Reconsider, now, the Jones/Jones* [roll-back] case. Entertain once again, the thought experiment that God has a thousand times caused the world to revert to precisely its state at the moment just before Jones decides to smoke, and that on about half these occasions, Jones decides to smoke and acts accordingly. Assume, again, that in each of the reruns, Jones was trying to do two competing tasks, and that whatever he ended up doing, he would have done voluntarily, intentionally, and rationally. Suppose, as I have proposed, responsibility gauges the moral worth of an agent with respect to some episode in her life – a person discloses what she stands for when she is morally responsible for some deed. Then, given *type identical pasts,* when Jones does one thing in half or so of the reruns but something else in the others, there is no saying what Jones stands

[89] In fact, even if I am wrong above, it does not seem to me that rationality matters that much. We do not typically excuse someone just because they have acted irrationally. Irrationality, in other words, is not normally taken to indicate an absence of responsibility grounding control. If someone is systematically irrational then we might be inclined to say that they fail to satisfy sanity conditions and for that reason are not morally responsible. But even here, this would be because the agent was incapable of acting rationally,

for. (2002, pp. 118-119)

The conception of responsibility this criticism presupposes is one according to which unless the decision that the agent makes in the actual sequence is strongly connected (for which we can read deterministically caused) by the agent's character, then the decision cannot be "deeply reflective of who we are" (Double 1996, p. 76). The agent is only morally responsible if their decision "discloses in conduct one's moral stance or commitment vis-à-vis a particular episode in one's life" (Haji 2002, p. 118). Haji claims that "[o]ne can't... disclose what one morally stands for with respect to a particular action without its being the case that there is a contrastive explanation, in terms of prior reasons, of why the agent performed that action rather than some other" (2002, p. 143).

Haji's conception of responsibility draws heavily on what Gary Watson famously referred to as responsibilities 'attributability' face. Watson uses a quote from Dewey to capture the notion he is after:

> when any result has been foreseen and adopted as a foreseen, such result is the outcome not of any external circumstances, not of mere desires and impulses, but of the agent's conception of his own end. Now because the result thus flows from the agent's own conception of an end, he feels himself responsible for it... The result is simply an expression of himself; a manifestation of what he would have himself to be. Responsibility is thus one aspect of the identity of character and conduct. *We are responsible for our conduct because that conduct is ourselves objectified in actions.* (Watson 1996. P. 227)

Watson goes on to say that "[t]he self-disclosure view describes a core notion of responsibility that is central to ethical life and ethical appraisal. In virtue of the capacities identified by the self-disclosure view, conduct can be attributable or imputable to an individual as its agent and is open to appraisal that is therefore appraisal of the individual as an adopter of ends. Attributability in this sense *is* a kind of responsibility" (1996, p. 229).

---

so the problem seems to be one to do with determinism and the absence of relevant alternative possibilities.

Haji takes the self-disclosure view to imply that unless the agent's acts are strongly connected to their prior reasons, meaning *deterministically caused* by prior reasons, then the agent's act fails to disclose what the agent stands for. As such the agent is not responsible.[90]

What is assumed is that we are 'set' so to speak, and our decisions indicate the kind of person – our character values etc – that we hold.[91] Here attributability is achieved because our decisions can be sourced to some concrete prior 'us' that is revealed or expressed in our actions. But to this concern the first thing we can do is ask the compatibilist exactly how determinism can plausibly help achieve the kind of tight connection between an agent's action and their character in the kinds of cases we are talking about. For we need to remember that the cases we are talking about – the cases where the agent has genuine alternative possibilities – are ones where the agent is torn between different courses of action. It strikes me as implausible that a deterministically resolved torn decision can be said to be more expressive of what the agent stands for, than an indeterministically resolved one. In both cases the decision will reveal something about the agent, and in both cases the decision can reveal something equally enduring.

There is a related point. How exactly does the internal presence of indeterminism prevent this kind of evaluation from being made? It is true that the indeterminism present in directly free acts means that the agent's past character does not entail that they will make one decision or another. But the agent's decision sets their will (at least until

---

[90] Haji's view here can be compared in relevant respects with that of Hume, for whom an agent can only be responsible if their acts are attributable or revealing of their character.

> Actions are, by their very nature, temporary and perishing; and where they proceed not from some *cause* in the character and disposition of the person who performed them, they can neither redound to his honor, if good; nor infamy, if evil. [A] person is not answerable for [immoral acts] if they proceeded from nothing in him, that is durable and constant. (Hume 1955, p. 98)

The clause that a person is not answerable if their action "proceeded from nothing in him, that is durable and constant" does imply that an agent cannot be responsible for acts that are out of character – in this respect Hume's view is surely wrong, for intuitively we *are* responsible for such acts so long as certain control conditions are satisfied. The constancy condition would, plausibly, not be met in the case of indeterministically resolved torn decisions – but as I point out later, this is an unattractive implication, for it may turn out that our torn decisions really are indeterministically resolved.

[91] When I say 'set' I do not mean to claim that such a view presupposes that we cannot change over time. It is just that in whatever way our characters evolved, this was a function of how they were, combined with environmental factors. This kind of 'changeability' contributes nothing to 'ownership'. The compatibilist would have to say that we have as much 'ownership' over what we do even if, as it turns out, our characters and all the determinants of our decisions *are* fixed and unchanging.

they make another torn decision) in which case we can say that the agent's decision does indicate how they are – it reveals what they've just made themselves, rather than revealing what they were that led to them making the decision. As far as I'm concerned this achieves everything in terms of attributability available in deterministic settings. The agent's indeterministically resolved torn decision does reveal what the person stands for *now*.

What we need to recognise is that on Kane's modest incompatibilist view our will is not 'set', by our character and values– we find ourselves incomplete insofar as we are regularly subject to SFAs, and it is through resolving our torn decisions that we take a stance. It is as a result of this on-going self-creation process that part of what we are becomes non-attributable to factors external to ourselves, and so deeply attributable to oneself. This is something that I am going to say more about in the following chapter, but at the moment it is enough to note that there is a strong case for saying that it is precisely through agent-internal indeterminism that one gains the depth of attributability needed for responsibility.

6.8 Un-attractive Implications

If I have been wrong so far and indeterminism, as located on the Kanean view, erodes control in a responsibility-subverting way, or in some other way (rationality, attributability) works to subvert responsibility, this would land Haji, and any other compatibilists making these criticisms, with some unattractive implications. For the fact is, that indeterminacy really might be a feature of our torn decisions. Kane's account is quite consistent with all that we currently know about the brain and its operations – the brain is a dual processor, and it is possible that there could arise tension between the two processes, boiling up indeterminism and so forth – and these occasions could be experienced by us as our torn decisions. If this does turn out to be the case Haji would find himself having to insist that we are not responsible for our torn decisions, though we *are* responsible for our non-torn decisions (by hypothesis, the indeterminacy in the will is not 'stirred up' on these occasions). This is rather peculiar. There must be something wrong with a view which says this. Either responsibility goes or it stays; a plausible view should not allow it to hang around in scraps like this.

6.9 Conclusion

In this chapter I have argued that agent-internal indeterminism does not erode compatibilist control in a way that could plausibly subvert an agent's responsibility. In the case of Valerian incompatibilist views, the indeterminism does not erode the agent's active control in any way. I allowed that certain varieties of non-active control might be eroded to some extent, but that erosion in such domains could not plausibly be taken to subvert responsibility, and that determinism does not guarantee any greater control.

With respect to the Kanean modest incompatibilist position the indeterminism is internal to the agent's will on the occasion of torn decisions. I argued that when one compares deterministically resolved torn decisions with indeterministically resolved ones, there was no reason to see the indeterminism as eroding control in any responsibility subversive way.

I then went on to consider other ways agent-internal indeterminism might be seen as threatening, including threatening the rationality of the agent's choices, and the attributability of the agent's choices. The concerns ere, in all these cases, misguided or question begging.

I do not pretend that any of the arguments that I have given are decisive. Throughout this chapter I have argued as if I am addressing a compatibilists audience. A control-incompatibilist would deny that we would be responsible irrespective of whether our decision making processes are deterministic or indeterministic so long as in the latter case the indeterminism is left unsupplemented. However, even the control incompatibilist could agree that strictly speaking the indeterminism, as it is introduced by either Valerian or Kanean incompatibilist, would not render us more non-responsible than would be the case if determinism were true.

I want to end this chapter with a concern to be taken up in the next chapter. Assume that my arguments above go through. When one combines the conclusion of this chapter with the previous chapter, one might be left wondering what value there is in agent-internal indeterminism. For whilst it may be reassuring that such indeterminism does not do us harm in terms of control, if it does not do us any good either, then it is reasonable

to wonder why it matters whether such indeterminism obtains or not. Why not just be a compatibilist? What positive work is the indeterminism doing? How is it paying its way? Following Clarke, we can call this the 'gratuity' objection (2002, p. 374). It is this gratuity objection that I take up in the next chapter. There I argue that agent-internal indeterminism can provide us with the kind of independence and ownership over our acts and characters that we saw CNC controllers threaten in Chapter 4.

Chapter 7

The Helpfulness of Indeterminism

I ended the previous chapter with a problem: if agent-internal indeterminism neither enhances nor erodes control, what positive work is it doing? In this chapter I shall argue that if we have genuine alternative possibilities this provides us with a kind of ownership unavailable under determinism.

I will argue that it is important where in the agent's deliberative process the indeterminism is located. I will reject the Valerian model that places the indeterminism early in the deliberative process, defending instead Kane's modest incompatibilist position. I will consider a number of criticisms and refine modest incompatibilism as necessary.

I will end by arguing that we should assess modest incompatibilism by how well it fares against rival views, and that certain dialectical delicacies need to be taken into account. Whatever one's view about the compatibility of responsibility and determinism, modest incompatibilism is, I argue, a more attractive view and provides moral responsibility with a more robust grounding than its compatibilist rivals. But, I will acknowledge that my arguments so far will have little persuasive power against a committed control incompatibilist. Different arguments are required to do that, and they will be supplied in the next chapter.

7.1 Ultimacy and Ownership

The conclusion from what has been argued so far is that the only kind of control we have over what we do, is some form of compatibilist control of the reason responsive variety. But, control is not plausibly all there is to having the kind of free will required for moral responsibility.[92] As noted in previous chapters, part of the intuitive picture of

[92] There are conditions that are strictly speaking neutral between compatibilism and incompatibilism, such as sanity and epistemic conditions. It is because of the neutrality of such conditions that I have ignored them in this work. That is obviously not to say that they are not important conditions that a full account of free will needs to discuss. I have chosen instead to focus on control and ownership insofar as these issues divide the compatibilist and incompatibilist, and to see how these issues relate to a wider problem with

free will is to be the ultimate source of one's actions in some sense. It is here that we find value in indeterminism. For if indeterminism is located in such a way as to generate genuine alternative possibilities, we can be said to gain a valuable kind of independence from factors external to our conscious wills, and in that way indeterminism can provide us with sourcehood and thereby a degree of ownership over what we do. For to some extent what we do, and become, is rendered ultimately down to us.

There's a compatibilist sense in which we are the sources of our actions – for it is still we who make our decisions, even if determinism is true. Trying to put a positive gloss on this compatibilist sense of sourcehood, Fischer says "even if there is just one available path into the future – I may be held accountable for *how I walk down this path*... Even if I somehow discovered there is but one path into the future, I would still care deeply how I walk down this path... I would want to do it my way" (1994, p. 216). Fischer later refers to this alternative conception of ownership as associated with what he calls "the importance of indispensability" (2003, p. 207).

> Note that even if causal determinism obtains, invocation of prior states of the world plus the natural laws cannot explain our behaviour and its upshots without *also* explaining that *we make a certain sort of contribution to them.* That is, the prior conditions and laws of nature explain what happens only by also explaining that we make a certain sort of contribution – that or deliberations have a certain character, for example. (2003, p. 207)

However, the problem with these compatibilist conceptions of ownership or sourcehood is one that we are already aware of from Chapter 4, namely that we can satisfy such conditions *whilst being CNC controlled.* Fischer does try to rule out CNC control by saying that indispensability involves our making a contribution through our "unhindered deliberations" and then says that by unhindered he means "deliberations not impaired by factors *uncontroversially* thought to rule out moral responsibility" which includes, conveniently, "manipulation" (2003, pp. 207-208). But again here we have Fischer just ruling out CNC manipulation by stipulation, just as we saw him doing in Chapter 4. For

---

luck that seems to be faced by *both* positions. For an excellent discussion of free will and 'sanity' see Wolf (2003).

in what way exactly does CNC manipulation hinder or impair one's deliberations *except* by antecedently determining their course?

This susceptibility to CNC control is precisely what brings out the lack of depth to compatibilists accounts of ownership, and why they cannot plausibly satisfy any requirement of ownership associated with free will and moral responsibility. Although ownership is, admittedly, a rather vague notion, once we accept that the only kind of control on the market is compatibilist control, then we can gauge the value we attach to independence and ownership in direct proportion to how much we disvalue being subject to CNC control. We can gauge how successfully a view provides us with ownership over what we do by how successfully its conditions for free will thwart CNC controllers. In other words, a view on free will provides us with the depth of ownership necessary for responsibility if it insulates us satisfactorily against CNC control. In this way, despite the unavoidable vagueness of ownership we can accurately test whether a view provides it.

7.2 Ownership and CNC Control

We saw in Chapter 4 that a sophisticated enough CNC controller could programme an agent to do what they wanted, consistent with the agent satisfying compatibilist free will conditions. The moral of such cases seems to be an incompatibilist one. For if free will entails a requirement for genuine alternative possibilities, then it would be impossible for a CNC controller to both design a free agent *and* ensure that their agent freely made one decision rather than another. For the CNC controller would have to design their agent precisely so that on certain occasions, it was genuinely open which decision the agent would make.[93] On such occasions the CNC controller could not reliably ensure that the agent made one decision rather than another. The most a CNC controller could do is to restrict an agent's option ranges or the frequency at which such alternative possibility moments turn up. Even then, as we shall see later, free will of the responsibility grounding kind also places restrictions on the extent to which a CNC controller could exercise this restricted form of control. However, whilst I think that we certainly should draw an incompatibilist moral from CNC control cases, it nevertheless

[93] The controller could bypass the agent's operating mechanism and in that way get the agent to make the decision that they want, but in that case the controller would be exercising constraining control and so

matters where in an agent's deliberative process the indeterminism is located.

7.3 Valerian Incompatibilism

We are, by now, familiar with Valerian incompatibilism. We can note that, by introducing indeterminacy early in the deliberative process, so that during the process of deliberation it is causally open (at least on a suitable number of occasions) which beliefs will come to mind it can be genuinely open what decision an agent's deliberative process will issue in. Introducing indeterminacy in this way does seem to secure genuine alternative possibilities. And as we saw in Chapter 5, Mele thinks that this is enough to secure the relevant kind of ultimate sourcehood for ownership.

However I hold that Valerian incompatibilist positions are no better at securing ownership of the valuable kind, than are straightforward compatibilists conceptions of free will. For a CNC controller could still subvert an agent's responsibility by implanting a 'Valerian' incompatibilist mechanism. In other words, a CNC controller could control an agent in an intuitively responsibility-subverting way, consistent with the agent satisfying Valerian conditions on free will.

Consider first that it is the fact that with genuine alternative possibilities the CNC controller is prevented from antecedently ensuring that the agent makes one decision rather than another that effectively thwarts their ability to control the agent. But there are different ways in which an agent might be unreliable with respect to decision-making. On the one hand an agent might be unreliable when it comes to making one decision rather than another. But it might be the case that the agent cannot be relied upon to make a particular decision, because they cannot be relied upon to stay alive, or not to have a stroke or such like. I think that it is fairly clear that this kind of unreliability does not secure the kind of independence or ultimacy needed for ownership.

For instance if the CNC manipulator had implanted a compatibilist control mechanism, but had also designed the agent so that it is indeterministic whether the agent remains conscious at certain point, this indeterminacy would not do anything to deliver ownership to the agent. For the fact is that whilst the CNC controller could not, due to

---

would cease to be a CNC controller and would violate even compatibilist free will conditions.

the indeterminacy, ensure that the agent made a decision at the appropriate time, they could ensure that if the agent made a decision, that it would be the decision that they wanted the agent to make. In this kind of case the agent's responsibility and free will is intuitively undermined because if they make any decision at all, they will only make that decision which the CNC controller programmed them to make.

In the above case the indeterminacy is wholly external to the agent's deliberative process. But are matters different where it is indeterministic which beliefs will come to mind? Well, indeterminacy at this point does seem to secure alternative possibilities in which the agent decides otherwise (as opposed to pathways in which the agent is dead due to a stroke). But it remains the case that the agent reacts in the way that he does to the beliefs that do come to mind because of the way the CNC controller designed them. In other words, the agent is a slave to what beliefs come to mind, and the form that slavery takes is a matter that the CNC controller will have determined. Although the agent does seem to have been provided with genuine alternative possibilities in which they decide otherwise, the agent still seems to lack the crucial kind of independence needed for ownership.[94]

7.4 Valerian Incompatibilism and Frankfurt-Style Cases

The following argument can also be added to the case against Valerian incompatibilism. In Chapter 4, I argued that a favourite tool of the compatibilists – the Frankfurt-style case - can actually be used to highlight that there is no relevant difference between being subject to CNC control, and being antecedently determined to act as one does.

What I now suggest, is that if an incompatibilist position facilitates the construction of a Frankfurt-style case, then this incompatibilist position, because it is capable of facilitating passive covert control, is therefore also capable of facilitating active covert control. In other words, one way of testing whether an incompatibilist view has located indeterminism in a place where it is capable of delivering independence and ownership,

[94] We can note something interesting here, something which underscores the way in which CNC controllers implicate determinism as the responsibility and free will subverting culprit. For consider that we can note from the above discussion that a reduction in the CNC controllers control, does not automatically yield an increase in the agent's ownership.

is to see if the indeterminism will rule out passive covert control of the kind exercised by Frankfurt-style counterfactual interveners.

Valerian incompatibilist positions fail this test. For by locating the indeterminacy early in the actional pathway, a Frankfurt-style counterfactual intervener is afforded a reliable basis upon which to intervene prior to decision-making. I will briefly outline such a case. Intervener Newman, wants Jerry to make a particular decision, P, at a particular time t2. Jerry's decision-making processes contain indeterminacy at the level of the coming to mind of certain beliefs. Newman is monitoring Jerry's brain, and knows that Jerry will make decision P at t2 unless a certain belief comes to mind at t1. It is indeterministic whether this belief will come to mind. However, if this belief comes to mind, then Newman will intervene in Jerry's decision making process, ensuring that Jerry makes decision P at t2. In the actual sequence the relevant belief does not come to mind, and Jerry makes decision P under his own steam. But there was no possibility of his deciding otherwise. And hence, PAP is refuted for Valerian incompatibilism. Valerian incompatibilism shows itself to be incapable of delivering the kind of independence needed for ownership.

7.5 Kanean Modest Incompatibilism

Whilst Valerian incompatibilism fails to introduce indeterminism in a way that could secure the kind of independence from the past required for ownership, I hold that matters are different when it comes to the kind of modest incompatibilist position developed by Kane. It is securing this kind of independence which is the primary motivation behind Kane's view.

By locating the indeterminism in the agent's effort of will on the occasion of torn decisions, it becomes indeterminate in what way the agent will consciously decide *up to the moment of choice*.[95]

[95] For reasons already made clear in Chapter 3, this kind of view does not run afoul of Frankfurt-style cases. For because the agent's effort of will is indeterminate up to the moment of decision-making, there is no prior basis upon which a counterfactual intervener can intervene. The view therefore passes my test above.

One reason I believe this delivers the kind of ownership we need for free will and responsibility goes as follows. If we did have incompatibilist control of the mysterious kind – so path-picking control of the kind that would eradicate agent internal causal luck - then we would have free will by just about anyone's standards (van Inwagen 2002, p. 168). And so, by hypothesis, we would have the requisite ownership over what we do. If, for instance, a CNC controller implanted us with a mechanism that (somehow) provided us with incompatibilist control, then our responsibility for what we do would be in no way threatened by this historical fact. For the CNC controller – as God is held by some to have done – would have *given us* free will. And thereby they would have relinquished control over us.

If we were to remove the incompatibilist control, whilst leaving in place the indeterminism accommodating it, then the CNC controller would have no greater control over what the agent did, than if the incompatibilist control were in place. In other words, the CNC controller is in no better position with respect to their control over what we do.[96] We would have as much independence from the past as if we have incompatibilist control. We would have as many paths into the future. In short, we would have as much 'up to us-ness' just through the will-internal indeterminism, as we would if such indeterminism were supplemented with incompatibilist control.

There are a number of concerns that might be raised against what I have just said. Firstly, there might be a concern that without the incompatibilist control in the picture, the agent would gain ownership, but at serious cost in terms of control. But this is to make the erosion argument and so the arguments of the previous chapter apply.

Secondly, there is a concern that in terms of independence and ownership one gets everything that one would get if one had incompatibilist control, but nevertheless maintains that without the addition of incompatibilist control one will not have enough control for responsibility, for responsibility requires something more than compatibilist

[96] One might object this would be also be the case if the indeterminism were located earlier, as with Valerian views. But this is not so – the fact that 'Valerian' incompatibilist conditions facilitate covert *passive* control by a counterfactual intervener are evidence of this. It actually *is* possible for someone to implant an incompatibilist mechanism and be able to ensure that the agent only ever makes certain decisions, whilst, in the actual sequence, it turns out that the agent acted freely on every occasion. In other words, on the Valerian view, we can have just one path into the future, but can travel that path freely, and all this due to the designs of a covert controller. None of this is possible with the Kanean incompatibilist position.

control (see, for instance, Clarke 2002, p. 376). This kind of concern returns us to worries about the adequacy of compatibilist control, and it is a concern that I will turn my attention to in the next chapter, rather than here.

Thirdly, there is a concern about whether the indeterminacy in the will really can thwart CNC control. It is this type of challenge that I will consider in what follows. I will consider a number of such charges below, including one that I raise myself, highlighting a novel way in which a CNC controller might be able to exercise their control over the 'Kanean' agent. Whilst I reject the other challenges to ownership, I argue that my own challenge calls for a slight modification in the modest incompatibilist view.

7.6 Objection: Haji

Haji agrees with me about Valerian incompatibilism. One is no better protected against ownership subverting CNC control with Valerian free will than one would be with compatibilist free will. But Haji thinks that if this is true, then the same holds true for modest incompatibilist positions of Kane's variety:[97]

> [I]f victims of [CNC control] are not responsible for their choices when indeterminacy in the actional pathway of events leading to choice occurs relatively *early* in the pathway, and they are not responsible because they are victims of [CNC control], how can they be responsible when indeterminacy occurs *further up* in the pathway ending in the choice when they are still victims? We can grant that in her conflict [the agent's] effort of will that culminates in her decision to smoke is indeterminate and her resulting choice undetermined. But presumably the mental actions that are an agent's efforts of will must, in some nontrivial way, depend causally on the agent's antecedent actional elements like her values, desires, and beliefs. But in [both modest-libertarian agent and robust modest-libertarian agent] the relevant antecedent actional elements have been heteronomously acquired. How, then, can [robust libertarian agent] be morally responsible for her choice to smoke if we grant

[97] Although Haji agrees with me about Meleian incompatibilism, I did not consider Haji's arguments in the preceding section. This was because Haji's argument was quite different to mine. Haji's argument was exactly the same as the argument that he has presented here regarding Kanean incompatibilism.

> that [modest libertarian agent] is not morally responsible for her choice to do A? (Haji 2002, p. 131)

So, Haji's argument is that if one agrees that indeterminacy early in the "actional pathway" offers nothing in terms of ownership, then moving the indeterminacy further along cannot plausible change matters. I disagree, for the move along the actional pathway is significant. The move means that indeterminacy is, as I've said above, now in the place where incompatibilist control would operate if we had it. The indeterminacy is now at the centre of the action.[98]

But Haji's more basic point seems to be that if all the basic elements of an agent's free will mechanism have been implanted, and if the decision that the agent makes is sourced to those elements, then the agent's responsibility is undermined and no amount of indeterminacy between those elements will make any difference.

But this assumes that the real problem that cases involving CNC manipulation brings to our attention is that the agent themselves did not do the job of implanting their actional elements. In other words, the real problem is that the agent did not "pull oneself up into existence by the hair, out of the swamps of nothingness" (Nietzsche 1966, p. 21). But this is an unduly demanding self-creation condition to saddle the incompatibilist with and I would reject that any plausible incompatibilist position can be committed to it. I am going to say more about this in the next chapter – for it is common for the critics of incompatibilism to invite us to associate the position with various outlandishly demanding conditions. But for my purposes here, I think the quickest way to see that ownership cannot involve 'pulling oneself out of the swamps of nothingness' is to once again return to my principle argument. If an agent had incompatibilist control they would, by hypothesis have sufficient ownership over what they do. But quite clearly this agent does not have unrestricted ownership. The choices that this agent makes and the character formation process that they undergo, will be made against a background not of the agent's choosing. To use a well-worn metaphor: the garden of forking paths the agent gets to negotiate is not one of the agent's making – but the agent *does* get to walk their own route within that garden. All of this stays the same if one removes the

incompatibilist control but leaves in place the indeterminism. For the incompatibilist control did not give one more alternative possibilities, it did not enlarge the scope of one's background character traits and so forth. The garden remains as much outside your control as it would do if you had incompatibilist control, and the route you take through that garden is as much down to you as it would be if you had incompatibilist control – for after all, nothing external to your will is going to determine which paths you take.

What Haji misses is that the problem covert control highlights is not strictly speaking the fact that one's mechanism has been implanted (for ultimately one never controls the process of mechanism acquisition, implanted or not), but the fact that all that one subsequently does is now determined. All that one subsequently does is explained fully in virtue of the fact that one had in place a mechanism of this or that kind in circumstances of this or that nature.[99] To return to the metaphor of the garden of forking paths: we do not need an endlessly large garden, nor do we need to have built the garden ourselves. It is enough that the garden in which we find ourselves contains open pathways of a certain quality; and it is enough that it is down to us with pathways we travel down. That is enough for us to have independence and ownership. It is a 'restricted' form of independence and ownership but, a restricted form of independence and ownership is *still* a form of ownership. And this form of ownership is *still* of a deeper kind than available under determinism.

---

[98] Note that Haji himself must acknowledge the significance of this move along the actional pathway, for we saw in the previous chapter how, according to Haji, this move yields an erosion in control, whereas, by his own admission, it would not if it were located earlier.

[99] Schlossberger also misses this point when he makes the following argument:

> Suppose that William decides to give me a particular set A of beliefs, attitudes, dispositions, etc. He performs neurosurgery, altering my brain so that I have set A. Now, by chance I happened to have set A before William began his procedure. Thus my brain after the procedure is no different than it was before William began. As a result, my subsequent actions are o different than they would have been without the procedure. I am just as susceptible to rational persuasion as I was before, etc. Does William's procedure make any difference to my moral status? Am I henceforth absolved of all moral responsibility? (1986, p. 45)

Schlossberger is a compatibilist, and he is using the above point to argue that alternative possibilities are not needed for responsibility. But at the moment let us just note that if Schlossberger's initial mechanism satisfied the modest incompatibilist requirements, and if the implanted mechanism *also* satisfied those requirements, then Schlossberger will be responsible for his subsequent decisions. But then it remains the case that he will have alternative possibilities (for this is just part and parcel of satisfying the modest libertarian condition).

7.7 Objection: Double

I will now turn to a criticism made by Richard Double. Double has argued that a CNC controller could still exercise a destructive kind of control over an agent, even if the agent did satisfy modest incompatibilist free will conditions, for "a [CNC] controller could set up undesirable alternatives between which we would make indeterministic choices that would leave us woefully unfree under the controller's disjunctive control" (1996, p. 139; 1991, pp. 214-215; 1989). In other words, a covert controller might implant a mechanism which is indeterminate in the right kind of way, yet the indeterminacy might range over a limited range of choices. For instance, a covert controller might want Jones to assassinate either the president or the first lady - either will do. Jones's reason-responsive mechanism has been implanted with this in mind, and although it will be indeterminate which way Jones will choose, it is nevertheless the case that the implantation ensures that he will choose one of these options. The argument would go that in this kind of case Jones's moral responsibility is intuitively undermined, and yet by my lights Jones had ownership over his decision.

In response the first thing I would do is concede that Jones did have ownership over his decision, because it was ultimately down to him, there and then, to decide whether to assassinate the first lady or the president. Nothing external to Jones made him decide as he did. Similarly, if Jones were insane and his option range consisted of which out of a range of songs to bellow at the top of his voice, then I would again grant that his decision would be one over which he'd have ownership. But this would be equally true if, in both cases, Jones had genuine incompatibilist control over which decision he made. For as already pointed out, such luck-eradicating control does not give the agent control over their option ranges, but only over which option they actualise from within that range. Having luck-eradicating control does not, for instance, give you a guarantee that you'll not go insane, nor does it guarantee you that a mad CNC controller might not restrict your option ranges, as in the case above. In both cases – going insane and having one's option range restricted – one has not been deprived of one's luck-eradicating control. Similarly, one has as much ownership over what one does when insane, or when one's option range is restricted, as in standard cases.

But does this mean that I am committed to the absurd claim that Jones is morally responsible for his decision to shoot the first lady rather than the president, or that insane Jones is morally responsible for his decision about which song to bellow? Yes and no. Imagine that the Jones in the above cases had incompatibilist control. Well, then he would be responsible, but he'd be responsible for 'deciding to shoot the first lady rather than the president'. In the insanity case, he'd be responsible for 'deciding to bellow song X, rather than song Y'. In other words, he'd be responsible for some contrastive fact. In both cases, the contrastive fact in question did not contain either a morally permissible option (in the case of the assassination) or a morally significant option (in the case of the song-bellowing). As such, these are not the kinds of decision for which the agent can be blameworthy or praiseworthy. In both cases the agent is 'responsible', but not morally responsible.

What we can say is that where the kind of free will needed for responsibility is concerned, it is important not just that the agent have genuine alternative possibilities, but also that they are of a certain quality. If the decision that the agent makes is one for which she is to be morally responsible, it must be possible for the agent to have avoided wrongdoing, and thereby avoided blame (Harrison 2005; McKenna 1997, pp. 73-75; Wyma 1997, p. 59; Otsuka 1998, p. 688). Otsuka calls this the "Principle of Avoidable Blame":

> PAB: One is blameworthy for performing an act of a given type only if one could instead have behaved in a manner for which one would have been entirely blameless. (1998, p. 688)

Assuming that one can only have the possibility of behaving in a manner for which one is entirely blameless if behaving in that way is morally permissible, then PAB entails that responsibility requires an option range which includes morally permissible options.[100]

But it is important to recognise which way around my account puts matters. If we are to be the ultimate source of our act, then nothing external to us can have been a sufficient

[100] PAB, it seems to me, is an intuitive principle, and it is in part surely because of considerations to do with PAB that PAP gets much of its intuitive appeal (a point I make in my 2004).

cause of our act. If this condition is satisfied, then it will be the case that we will have genuine alternative possibilities in which we decide otherwise. But what we are the ultimate source of - in other words, what we have ownership over – will be a function of the alternative possibilities that were available to us at the time of decision making. So, the assassin has ownership over their contrastive decision to assassinate the first lady rather than the president.

What we do is assess the moral significance of this contrastive choice. Choosing to assassinate the first lady rather than the president is actually not that morally significant (unlike, for instance, choosing to assassinate the first lady rather than watering the plants). Where all alternatives are bad, then at most the agent can be responsible only for having picked the worst of two bad options. But where both are equally morally impermissible, then whichever choice is made, the agent's option range is not of a kind that can legitimise a judgement of moral responsibility – not because the agent did not have the right kind of control and not because the agent failed to be the ultimate source of what they did, but rather just because of the moral quality of what they did.[101]

The general point worth stressing can be put this way. We can say that it is important that we make a real difference, hence the need for alternative possibilities. It is also important that *we* make the difference, hence the need for internal indeterminism so located that the difference in question is ultimately sourced to 'us'. But it is also important that the difference we make be of a sufficiently significant kind.[102] For part of what responsibility marks is the fact of this significance. In cases where there was no

---

[101] One might raise some concern over my characterisation of an option range as being one that contains only morally impermissible options. For ought implies can, and ought-not implies can refrain, so to say that a particular option is morally impermissible is to presuppose that there is a permissible option available. Hence, option ranges in which all the options are morally impermissible are not possible.
This is exactly the same kind of argument that many would use to argue against the possibility of genuine moral dilemmas. An option range in which all the options are morally impermissible just *is* a moral dilemma, and so the arguments against moral dilemmas automatically apply. Now, let me say that I think the above point has merit, and would agree that in fact one cannot have such an option range – or at least, one can, but it would be incorrect to describe it as one in which all the options are morally impermissible. But this concession would not affect my point in any way. For moral responsibility presupposes wrongdoing, and as such option ranges in which no option is wrong cannot be ones for which the agent can earn blame or praise, precisely because the agent will not make a significant moral difference through their action.

[102] Another way in which the quality of the option range can be important concerns the sanity of the range in question. It is one thing to be torn between stopping and giving aid to someone in need, and walking past to attend the important meeting. But if one is torn between giving aid on the one hand, or screaming obscenities at the sky on the other then the agent is not responsible whichever way she chooses. For the option range in question is an insane one.

morally permissible option available, then whatever difference the agent makes, it will not be a morally significant difference. It will not matter how much control the agent exercised over making this difference.

7.8 Objection: Tracing Views

There is another way in which a CNC controller might still pose a problem for the modest incompatibilist account of free will. I do not think the problem insurmountable, and raise it only because I believe it prompts a refinement of the account.

To make this particular objection, it is necessary that we familiarise ourselves with tracing views. I have already briefly mentioned tracing views in the previous two chapters, but we now need to say more. In Chapter 5 we saw that Kane refers to genuinely indeterministic torn decisions as self forming acts or 'SFAs'. This is because on these occasions the agent themselves determines the course their character development will take. The idea is that the agent is, on these occasions, partly contributing to their own future character – building up a stake, if one likes, in their character. But Kane also holds that an agent does not only have free will on the occasion of SFAs. Kane holds that an agent can have decided freely even if their decision was determined by their character, so long as the agent had some stake in that character. Having such a stake would mean that there would need to be SFAs in the past – at some past time the agent would have had to have had genuine alternative possibilities. But genuine alternative possibilities are not always needed.

What we can do here is distinguish between directly free acts and indirectly free acts. We can say that directly free acts are SFAs. Indirectly free acts are acts which stem from reasons and motives that are themselves due to earlier directly free acts. In other words, indirectly free acts owe their 'freeness' to earlier SFAs. There we noted the intuitive attractiveness of tracing views: we consider the drunk responsible for their behaviour – at least to some extent – so long as we assume that they freely chose to get drunk.[103]

Kane is not alone in endorsing a tracing view. Aristotle endorsed the tracing view when

he said that "just as when one has once let go of a stone, it is too late to get it back – but the agent was responsible for throwing it, because the origin or the action was in himself. So too it was at first open to the unjust and licentious persons not to become such, and therefore they are voluntarily what they are" (1976, p. 124, 1114a 18-19). And it is common amongst contemporary incompatibilists – whether of modest variety, or control-incompatibilists.[104] For the alternative is to adopt what is known – following Fischer – as 'restrictivism' (1994, p. 47).

The restrictivist incompatibilist would restrict free will and responsibility only to those occasions where the agent had genuine alternative possibilities. There is disagreement amongst incompatibilists about just how often we do have genuinely available alternative possibilities. We saw in Chapter 5 that Kane outlined six kinds of occasion associated with SFAs, and thinks that we are subject to such occasions fairly regularly. Balaguar, another modest incompatibilist of Kane's stripe, thinks that we make torn decisions a great deal of the time, whilst van Inwagen thinks that we very rarely make such decisions. So, how restrictive restricitivism would be, is going to be a function of just how often one thinks we have genuine alternative possibilities. The fact is though, that whatever view one takes here, it is likely that restricitivism would lead to greatly diminishing what we are responsible for. A great deal of what we do, a great deal of the decisions that we make, are made unthinkingly, on the spur of the moment. For all such decisions, we bear no responsibility on the restrictivist view. If one thinks – as van Inwagen does – that we rarely have genuine alternative possibilities – then one is going to find oneself coming perilously close to being, to all intents and purposes, a hard determinist. So let us assume that a viable incompatibilist view (whether of modest or control varieties), or at least one that does not want to be too revisionary, will endorse the tracing view.

---

[103] Aristotle discusses such cases, pointing out that "penalties are doubled for committing an offence in a state of drunkenness, because the source of the action lay in the agent himself: he was capable of not getting drunk, and his drunkenness was the case of his ignorance" (1976, p. 123).

[104] Note that tracing views are not only held by incompatibilists. Compatibilists hold them too – see Fischer 1994, pp. 175-178. There is controversy over whether Aristotle himself was compatibilists or incompatibilist, Sorabji defending the incompatibilist interpretation in his 1980, but others defending a compatibilists interpretation (Irwin 1980; Fine 1981; Broadie 1991). It is arguably that any sensible view on free will must incorporate a tracing view of some kind. The problem that I draw attention to, is one that concerns incompatibilists only however, insofar as incompatibilists have a putative advantage on the issue of ownership and CNC control.

The problem with the tracing view can be brought out by some remarks of Aristotle's. Aristotle points out that we do not control our dispositions – our character traits – in the same way that we control our actions. Our dispositions are only directly controlled by us at the beginning "the individual stages of their development, as in the case of illness, are unnoticeable" (Aristotle 1976, p. 126). In other words, whether our actions actually lead to the development of a certain disposition is not something that is *up to us,* though it was *up to us* to have performed the acts which, may or may not, lead to a certain disposition being acquired. But if, through acting unjustly, a man develops a disposition to be unjust, Aristotle allows that "it does not follow that he can stop being unjust, and be just if he wants to – no more than a sick man can become healthy, even though (it may be) his sickness is voluntary, being the result of incontinent living and disobeying his doctors" (1976, p. 124). Aristotle does not think this precludes responsibility, precisely because it was ultimately down to the agent to have made themselves that way even though "it is no longer open to them not to be such" (1976, p. 124). But I am less sure about this, and I think most incompatibilists should be too. Imagine that, up to now, most of us have experienced a significant number of SFAs. That means that, as of now, our characters are substantially, our own creation. And now imagine that the universe has, at this moment, become deterministic. So, from now on, everything we do will be determined by the past and the laws of nature. And, as it happens, this means that our characters are now fixed (this is not a necessary implication of determinism – but I am just asking that we imagine that, in fact, this is the case in this instance). So, from now on it is not up to us to be different from how we now are. On the tracing view, we will still be responsible for our decisions, or at least for those decisions we make which trace to the pre-determinism SFAs.

That might not sound too worrying to begin with. But consider that a common worry about determinism, is that if determinism is true, our characters are fixed. If I am selfish and unkind I am destined to remain so. Compatibilists are quick to point out that this does not follow. Determinism is the thesis that there is only one future, but it does not follow from this that one has a "fixed personal nature" (Taylor and Dennett 2002, p. 271). One's fixed future may be one in which one's character undergoes many radical changes.

> The latter [a fixed nature] is cause for dismay, perhaps, but not the former [a fixed personal future] for it could very well be one's fixed personal future to be blessed with a protean nature, highly responsive to the "activity of the self." The total set of personal futures, "fixed" or not, contains all sorts of agreeable scenarios, including victories over adversity, subjugations of weakness, reformations of character, even changes of luck. It could be just as determined a fact that you *can* teach an old dog new tricks as that you can't. (Taylor and Dennett 2002, p. 271)

But note that this reply by the compatibilist fails to address the real concern. For whilst one's nature might be determined to be changeable, it might be determined to be fixed. Determinism does not preclude having a character that undergoes many changes during its history, but nor does it guarantee it. For a compatibilist this makes no odds to one's moral responsibility. Provided the fixed nature that you have is one that affords you compatibilist control over what you do – as well it might – then you are free and morally responsible.

At first glance the modest incompatibilist view seems to be tailor made to answer these kinds of worries. The modest libertarian can point out that on their view you will only be morally responsible for what you do, if your nature is not fixed. You will only be morally responsible if, at various junctures, it is genuinely possible for you to turn out differently. But this attractiveness is shown to be threatened by my comments above regarding the tracing view. We can bring this out by employing a CNC controller.

Imagine that a Frankfurt-style counterfactual intervener is monitoring Jones's decision making processes. Jones makes a directly free decision or SFA. The counterfactual intervener knows that Jones has just 'set his will' such that, if he remains this way he will perform an act which the intervener wants him to perform, several years hence, other things being equal. But other things will only be equal if Jones does not undergo any further self-forming episodes. So what the intervener does is render Jones's decision-making processes deterministic from now on.[105] In other words, whilst Jones

[105] One might wonder exactly how the CNC controller could do this without actively intervening in a way that would subvert ownership. Well, one way would be to so arrange things that Jones will never face another torn decision – something the CNC controller knows how to do, given their supreme intelligence, knowledge and power.

himself sets his will, the intervener ensures that it remains set. All the subsequent decisions that Jones makes, provided basic compatibilist free will requirements are met, will be ones for which Jones will be truly responsible by the lights of the incompatibilist.[106] Yet the intervener ensured that Jones would face no further opportunities to change his will.

Maybe in the short-term the intervener's intervention does not subvert the agent's responsibility. But in the long term it seems unfair that the agent should continue to be responsible, for (unbeknownst to them) they have no opportunity to change the way that they are. My intuition is that this form of CNC control, a form which involves depriving a partly self-owned agent from having any further opportunities to change themselves *is* subversive of responsibility. It is not immediately subversive of responsibility, but becomes so the longer the agent goes without having any chance to change themselves, or endorse/re-affirm the way that they are.[107]

I do not think that my criticism above is devastating to the tracing view, or modest incompatibilism.[108] Rather it highlights the need for a refinement. Where Kane focuses on past SFAs, I hold that what I have above shows that future SFAs also need to be made part of what it is to have free will. Free will is not just about having *had* genuine

---

[106] I assume that modest incompatibilists hold that it is enough that an agent satisfy compatibilist control conditions for indirect free acts. It may be that something less than compatibilist conditions are all that are needed. After all, in the 'drunk' example used earlier, the agent *does not* satisfy conditions of moderate reason responsiveness for their acts performed whilst drunk, yet it remains intuitive to consider them morally responsible for such acts.

[107] Kane does seem willing to accept that one could, at least, be eternally praiseworthy. For in a footnote Kane endorses a particular religious use to which his account of free will has been put by the philosopher James Sennett. Sennett argues that there is a problem of heavenly free will. For assuming that in heaven there will, of necessity, be no wrongdoing, then there's a problem saying exactly how those of us who get-in, will retain our free will. A compatibilist would have no trouble answering this question. But the problem then is that compatibilist free will doesn't facilitate the free will defence against the problem of evil. There isn't time here to explore this issue, but we can take it as read that *most* theists are libertarians. Sennett's suggestion is that those in heaven can retain their free will and responsibility provided the agents' character was created through a sufficient number of self-forming acts during their tenure on earth (Sennett 1999). So, in heaven the agent's acts are free and responsible because they trace to earlier, directly free acts on earth. Kane agrees that though he does not "usually engage in heavenly speculation" that "creatures in an orthodox heaven, if they acted at all, would continue to act "of their own free wills" in the sense of "wills of their own free making" (2002, p. 408 fn. 5).
However, I think that most incompatibilists – including Kane – would not be comfortable with the same implications regarding hell. Consider that if an agent has, through a series of SFAs, come to acquire, and have a stake in, a set of vicious character traits, then in hell – assuming that this is where they end up – they could find themselves *eternally* blameworthy, providing determinism obtains in hell. This, I take it, would be out of the spirit of incompatibilism. The agent needs to have the opportunity to change their character.

[108] Note that the points I have made above apply to all who incompatibilists who employ tracing principles, including control incompatibilists.

alternative possibilities, it is also about having alternative possibilities in the future. To return to the metaphor of the garden of forking paths, we need to have regular forks both behind and in front of us. If we go too long without having encountered a fork, then we are in effect no longer in the garden of forking paths and our ownership over what we have become starts to diminish. The kind of ownership that SFAs deliver is leasehold not freehold.

I admit that this leaves matters vague. I have no theory that would tell us how often an agent needs SFAs or the rate of the deterioration of our ownership in our background character traits. I have to hand this kind of matter over to intuitive judgement. I do not think that this is too much of a problem however. For my comments above regarding ownership and the tracing view would apply not just to modest incompatibilists, but control incompatibilists too – at least those who subscribe to the tracing view. The problem I have highlighted above, and to which I offer only intuitive judgement as a solution, is one that is not, strictly speaking, due to the modesty of the modest incompatibilist position. Furthermore, the problem is also one for a compatibilist insofar as they too subscribe to tracing views. For instance, the man who gets drunk freely, but who, whilst drunk, behaves appallingly, does not, whilst drunk, exercise sufficient compatibilist control over their behaviour (the will not be sufficiently reason-responsive).

7.9 Luck and Ownership

If my arguments above are right the interesting upshot is that causal luck is actually required to deliver ownership. So, whilst everything we do is ultimately a matter of luck whether determinism or indeterminism is true, it nevertheless matters what kinds of luck we have been subject to. If everything we do is a matter of constitutive and circumstantial luck we lack ownership or deep attributability. For in that case all that we do traces to factors external to our conscious willings. But if there is some causal luck in the mix, and if the causal luck in question is internal to our wills, then we do have ownership, precisely because of luck's presence. Another way to put this would be to say that whether determinism or indeterminism is true, everything we do is a matter of luck, but in the latter case there is the possibility of some of what we do being a matter of 'our' luck.

Luck, as noted many times, is often seen as toxic to responsibility. But here we see that it has a positive side too. It can actually deliver something that many of us value. If we could have a luck-eradicating ability to decide otherwise, then that too would give us the said ownership. Causal luck is not *necessary* for ownership (if path-picking control is possible). But incompatibilist-control of the luck eradicating variety would give us ownership for the *same* reason, namely the presence of alternative possibilities securing independence from the past.

7.10 Help from Hard Incompatibilists and Hard Determinists

In this section and the next I will try to allay some fears about the relevance of ownership to moral responsibility. In this section I will look to the hard incompatibilists and the hard determinists for insight.

I have so far said little about the occupiers of these camps. Hard determinists are incompatibilists who believe that determinism actually obtains and so believe that we lack free will and are not morally responsible.[109] Hard incompatibilists differ only insofar as they acknowledge that indeterminism could well be true, but think that we still lack free will because it is unlikely that it obtains in the right kind of way (or that we lack the kind of extra factor needed for incompatibilist control).[110] The difference between the two is academic, and I will refer to both as hard incompatibilists from now on. For what unites both is the belief that we lack free will.

What is interesting about the hard incompatibilists is they often spend a great deal of time trying to persuade us that the absence of free will is not as bad as we might think.[111] And what is interesting about most libertarians (so, incompatibilists who believe that we do have free will) is that normally believe the opposite.

---

[109] One might hold that this view is refuted by what we know from quantum mechanics. According to standard interpretations the world is indeterministic at the micro level (there are alternative interpretations which keep matters deterministic, namely Bohm's hidden variables interpretation (Bohm 1984)). However at the macro level indeterminism might be cancelled out for all practical purposes. It certainly appears so for most events, and it could ell be the case for our mental events too.

[110] The term 'hard incompatibilist' was coined by Pereboom (1995).

[111] Not all say this. Smilansky, though not strictly speaking a hard incompatibilist (he believes that free will is impossible as opposed to possible but not actual), thinks that things would be sufficiently bad for there to be moral reason to maintain the illusion of free will (2000).

But let us suppose that the hard incompatibilist is right, and that a great deal of what we take to be at stake, actually isn't. Nevertheless, the hard incompatibilist – if they are not to slip into being a compatibilist – will still accept that *something* will change if we lack free will. Rather than disputing what the hard incompatibilist says, we should instead focus on that 'something'. And we should ask what kind of free will it would take to deliver that 'something'. When we do this we can, I hold, see the centrality of ownership and the plausibility of modest incompatibilism.

Consider first Ted Honderich. Honderich believes that our intuitive notion of free will – the garden of forking paths model – supports certain 'life hopes'. This involves a "sort of life-hope, whether about being an actress, surviving a battle, or whatever, is to have a hope best characterized... [as the] hope for an unfixed future... in which we are not creatures of our environment and our dispositional natures" (1988, vol. 2, p. 22). In addition, determinism also threatens an important sense of individuality:

> [Under Determinism] what [the person] did is *explained* by something that is not individual to, or peculiar to [the person]... [our actions can be explained] in such a way that it would follow that another person of like dispositions would in the same situation perform a like action. (Honderich 1988, vol. 2 p. 68)

Now we can turn to the modest incompatibilist conception of free will and see that it delivers the unfixed future supporting 'life hopes' of the kind Honderich mentions above, and it also delivers the important sense of individuality that determinism undercuts. It does this *without* bringing in extra factors. Nevertheless, we might speculate whether 'life hopes' and senses of individuality have anything to do with moral responsibility of the desert kind. I believe they do – that we can see our practices as being in part about respecting or marking our individuality and uniqueness. But there is, I think, another way of showing the importance of ownership. Here we can turn to another prominent hard incompatibilist – Derk Pereboom.

Pereboom accepts that acknowledging that we lack free will means having to give up "our ordinary view of ourselves as blameworthy for immoral actions and praiseworthy for those that are morally exemplary" (2002, p. 479). And Pereboom believes that this

would damage our life hopes, because "[i]t is not unreasonable to object that life-hopes involve an aspiration for praiseworthiness"(2002, p. 481). But he argues that we can separate out certain aspects of our life hopes from praise and blameworthiness:

> [A]chievement and life-hopes are not obviously connected to praiseworthiness in the way this objection supposes. If an agent hopes for success in some endeavor, and if she accomplishes what she hoped for, intuitively this outcome can be her achievement even if she is not praiseworthy for it – although the sense in which it is her achievement may be diminished. (2002, p. 481)

What we can note from the above is that Pereboom does acknowledge that *not* being praiseworthy (or blameworthy) means recognising that to some extent the sense in which what one does is one's achievement is diminished. The moral I draw from this is that free will is that in virtue of which one's sense of achievement is enhanced. And as I have argued above, if what we do is ultimately down to us, and cannot be attributed to factors external to our conscious willings, then we *do* have reason to see what we do as *our* achievement in a deeper sense than would be the case under determinism. As such modest incompatibilist free will does provide what is needed to render us apt targets for the reactive attitudes.

My argument here goes as follows. We first ask what exactly we would lose through the recognition that we are not truly praise or blameworthy for anything that we do. It is not unreasonable to draw Pereboom's conclusion that we would lose a certain sense of achievement. Of course, we might still be actively praised and blamed, for these practices might be justified on free-will independent grounds (their utility for instance). And we might still express blame and praise for others (again, on free-will independent grounds). But beneath it all, our recognition that we are not *truly* blameworthy or praiseworthy would rob us of that sense of achievement.

I believe that we should infer from this that part of what being truly praiseworthy or blameworthy is about is recognising our achievements (and failures). The recognition that determinism is true (or recognising that we are subject to CNC control) yields the same sense of loss of true achievement as comes from recognising that we are not truly praise or blameworthy for what we do. Now I think that it is not unreasonable to

conclude from this, that true blame and praiseworthiness is intimately connected with the (admittedly vague) sense of loss of achievement that comes through recognising that we are not the ultimate originators or sources of what we do. A conception of free will that restores that sense of achievement can therefore reasonably be seen as of a kind capable of grounding true blame and praiseworthiness.

In other words, our actions are ones for which we are aptly blamed or praised if they are attributable to us in a sufficiently deep sense. The kinds of achievements and failures that we make under determinism are not of the right kind, precisely because they are not ultimately attributable to us. We lack ownership over such acts. But if we have modest incompatibilist free will then our actions are more deeply attributable to us, and so we can see our acts as truly *our* achievements, and as such acts for which it would be apt to praise or blame us.

So, what Pereboom and Honderich do is help focus our attention on what work free will actually does, and why we might value being truly morally responsible. Both think that once we focus our attention in this way we will see that the absence of free will is not so bad after all (indeed, it may have benefits – see Pereboom 2002, pp. 487-488 and Smilansky 1994). If this is true then it actually takes the heat off free will to some extent. For to the extent that our practices and life hopes do *not* depend upon our possessing free will, then a conception of free will does not have to show how it can support such practices. What free will of the kind associated with moral responsibility has to do, is just make-good that which we would lose through not being truly morally responsible. If that comes down to losing a sense of achievement and true ownership over what we do, then *that* is what a plausible conception of free will has to restore. And that is precisely what modest incompatibilism does.

7.11 Responsibility and Dialectical Delicacies

Whilst sympathetic to the need for genuine alternative possibilities, and sympathetic to the need for the deeper kind of ownership that alternative possibilities provide, some are going to remain doubtful that the modest incompatibilist accounts improve prospects for moral responsibility over straightforward compatibilism. What I have just argued in the above section may go some way towards alleviating such concerns. But I suspect that

they will still harbour doubts that in whatever sense we gain a deeper sense of achievement with modest incompatibilist free will, the fact remains that we still fall short of having what it takes to ground true responsibility. For instance consider Clarke's remarks.

> I find less reason to think that an event causal-causal libertarian view improves upon a comparable compatibilist account when it comes to moral responsibility. If determinism is true, misdeeds may still be wilful and deliberate, exercises of a capacity to consider reasons and act on one's appreciation of them. There may consequently still be a type of justification for praise and blame, reward and punishment. Such reactions may be appropriate expressions of our feelings and our judgements about past behaviour; they may contribute to moral education and may encourage good behaviour and discourage bad; and they may help protect us from miscreants. Incompatibilists typically hold, however, that a very important type of justification would always be lacking: none of these reactions would ever be deserved. Less categorically, some incompatibilists allow that, in a deterministic world, there might be a type of desert of these kinds of reactions or of some version of these reactions, or that they may be deserved to a degree. But it is then said that there would be an important type of desert missing, or that an important version of these reactions would not be deserved, or that these reactions would not be fully deserved... Whatever the implications of determinism for desert really are, the implications for desert of our having just the variety of active control that is characterised by an event-causal libertarian view are, it seems, the same. (2002, p. 376)

Clarke is a control incompatibilist. To a control incompatibilist the modest incompatibilist position sound *half* right. By the control incompatibilist's lights what modest incompatibilist positions get right is the need for genuine alternative possibilities, and the value of the kind of ownership that such alternative possibilities deliver. But a control incompatibilist is normally convinced that the main problem with compatibilism is the inadequacy of compatibilist control. By a control incompatibilist's lights, compatibilist control is just not up to the job of making us truly morally responsible. From the control incompatibilist perspective therefore, the modest

incompatibilist view will appear to fall short of giving us all that is needed. If a view tries to get away without enhancing our compatibilist control, then the control incompatibilist is going to find it wanting. This is exactly the source of Clarke's concern:

> [The modest incompatibilist view] secures a type of leeway or openness not available in a deterministic world, but the view provides the agent with no additional positive power to determine what he does; it does not secure any greater degree of active control. And this is what seems to be needed if there is to be a different verdict concerning desert and hence responsibility. (2002, p. 376 note omitted)

A control incompatibilist, whilst almost certainly sympathetic to the need for a deeper kind of ownership than available under determinism is probably not going to be taken all the way to modest incompatibilism by what I have argued so far, because after all, their own view delivers the deeper kind of ownership too (for alternative possibilities of the kind that deliver ownership are required to accommodate incompatibilist control).

In Chapter 5 I argued that incompatibilist control is either incoherent or inherently mysterious. That is certainly something to be counted against control incompatibilism and which does not similarly apply to modest incompatibilism. But those were very common charges and the control incompatibilist is used to them. The fact is that the control incompatibilist is willing to bear a degree of mystery as an acceptable cost of their position. So, to persuade a committed control incompatibilist something more than what I have argued so far is going to be needed. I believe such arguments are available, and will present one in the next chapter.

We should not assess modest incompatibilism solely by how attractive it appears compared to control incompatibilism. We should also assess it by how attractive it appears compared to compatibilism. And I certainly hold that what I have argued so far should have bite against the compatibilist. CNC cases present a serious challenge to the viability of compatibilism, one that compatibilists themselves are sensitive to (see Watson 1987). As I have shown above, the modest incompatibilist position offers a principled way around the problem. I have argued in previous chapters that an important

part of the modern compatibilists armoury, the Frankfurt-style case, does not work. This means that in addition to the concerns over CNC control the compatibilist has stacked against them the fact that they can neither satisfy the intuitively plausible PAP condition, nor reject it on any non question begging grounds.[112] The modest incompatibilist position, on the other hand, validates PAP and provides us with free will of a sort impervious to CNC control.

It is true that the compatibilist position is a simpler position than the modest incompatibilist position – but this simplicity comes at considerable cost, as we have seen, costs that I believe it to be foolhardy to bear. It is also true that the compatibilist has the luxury of not having to worry so much about discoveries regarding the structural underpinnings of the universe (on this see Fischer 1994, pp. 6-7 & p. 207). But this in itself is not a good reason to be a compatibilist. A view does not become more likely to be correct, simply because the conditions it maintains need to obtain are likely to obtain. After all, imagine that, as a result of some disturbing discoveries, it becomes highly likely that we are being CNC controlled by super-intelligent beings on another planet. This discovery would not in itself increase the plausibility of the hard compatibilist position (the hard compatibilist being one who believes free will and CNC control are compatible). So we should not judge compatibilism by how secure it renders its conception of free will compared to modest incompatibilism. If we do, then the least demanding position is always going to win. And that is preposterous.

So, when compared to compatibilism, I hold that my arguments so far show modest incompatibilism to be a very much more attractive position. And so when Clarke says that "[w]hatever the implications of determinism for desert really are, the implications for desert of our having just the variety of active control that is characterised by an event-causal libertarian view are, it seems, the same" he is very much mistaken (2002, p. 376). His mistake stems from the fact that he is assessing the modest incompatibilist position from the perspective of a control incompatibilist and thereby fails to attend to certain dialectical delicacies (2002, p. 376). We should judge modest incompatibilism by how well it fares against its rivals. My arguments so far should be seen as primarily highlighting the advantages of modest incompatibilism over compatibilism. And on that

[112] Another advantage, one that I have chosen not to pursue here, is that through validating PAP, the modest incompatibilist view also validates the Kantian principle that ought implies can. It is a cost of a

front I believe that as a view about the kind of free will needed for moral responsibility modest incompatibilism wins resoundingly over compatibilism.

7.12 Conclusion

In this chapter I have argued that agent internal indeterminism can provide us with a degree of independence from the past and thereby provide us with a deeper kind of ownership over what we do than available under determinism. The shallowness of compatibilist 'ownership' was brought out in Chapter 4 by cases involving CNC control. Above I have argued that agent internal indeterminism can provide insulation against CNC control, and thereby secure the depth of ownership required for moral responsibility.

It matters where indeterminism is located, and I have argued that if it is located too early in the deliberative process – namely in the coming-to-mind of beliefs and desires – then it will not contribute to an agent's ownership. However, if, as with Kane's account, the indeterminacy is between the agent's reasons and their decision – in their effort of will in other words – then it can deliver the kind of ownership in question. For nothing external to the agent's effort will explain why they decided as they did. Indeterminacy here makes the agent no more or less an ultimate source of their action, than if they had incompatibilist control.

Whilst we might be able to gain broad agreement on the intuitive value of independence and ownership, there remains a persistent concern that without the supplement of incompatibilist-control, then whatever ownership we may gain through indeterminism, it is not enough to take us all the way to true responsibility. In response to this concern I argued that we should assess modest incompatibilism by how well it compares to rival views, and when we do this we need to respect certain dialectical delicacies. For instance, when judging how well modest incompatibilism fares against compatibilism, we should not assume the perspective of a control incompatibilist, as this will distort the assessment.

---

compatibilist view that it has to reject the Kantian principle (see Fischer 1999a; Haji 2002).

I ended by claiming that my arguments so far show that there is a strong case for favouring modest incompatibilism over compatibilism when it comes to moral responsibility. But I also acknowledged that I have said little that would persuade a control incompatibilist to favour modest incompatibilism. That is the task I turn to in the next chapter.

Chapter 8

Taking The Sting Out of Luck

In Chapter 5 I argued that short of appeals to mysterious extra factors everything an agent does is ultimately a matter of luck, irrespective of whether determinism or indeterminism is true, for the only kind of control that we have is compatibilist. In the previous chapter I argued that the agent-internal indeterminism characteristic of Kane's modest incompatibilism can provide us with an independence yielding a deeper kind of ownership over what we do than available under determinism. In other words, it can matter how we are subject to luck. A certain kind of luck – agent-internal causal luck – can make it the case that what we do, and what we become, is to some degree 'up to us' rather than being solely attributable to factors external to us.

Nevertheless, however necessary independence and ownership may be for responsibility grounding free will, the exposure to ultimate luck will, for many, be the crux of the issue. Many will maintain that whatever value attaches to ownership, the issue of ultimate luck settles the issue of free will and responsibility in favour of nihilism (Smilansky 2000; G. Strawson 2002; 2003; Pereboom 2002; Clarke 2002). Extra factors are needed and, if not available, there is no alternative to drawing the hard incompatibilist conclusion: none of us are morally responsible for anything.

But I believe that for those not already predisposed to nihilism about moral responsibility, such a conclusion is premature. Those predisposed to introduce extra factors, may yet be persuaded they are not needed. In this chapter I will argue that there is a strong case for believing ultimate luck is not the problem many take it to be.

First I will consider Fischer's attempt to deal with ultimate luck. I agree with Fischer's basic approach, but will use it to call into question the rationale behind the incompatibilist demand for path-picking control.

I will claim it is uncontroversial that if we could have incompatibilist control we would satisfy control requirements for responsibility. I will then warn against identifying incompatibilist control with a demand for unrestricted control.

I will go on to argue that once we recognise that incompatibilist control, if we could have it, would only amount to a restricted form of control, we must recognise the widespread and ineliminable involvement of luck in all that we do and its consistency with our moral responsibility. If we insist that the reason incompatibilist control makes us uncontroversially responsible is due to its luck-eradicating properties, then we are committed to what I term 'anti-luckism'. If one is to be true to one's anti-luckism then the remaining influence of luck is in need of neutralisation. Following Zimmerman, I will outline just what such neutralisation would imply.

I argue that on reflection it is sensible to drop the anti-luckist rationale for incompatibilist control. It is not plausible that incompatibilist control is needed for luck eradication purposes: it performs so poorly on this front and commits us to radical revisions to our practices. This gives us reason to explore the possibility that incompatibilist control could be associated with the satisfaction of some *other* condition – one consistent with ultimate luck. The requirement for independence and ownership discussed in the previous chapter is the natural candidate.

Finally I argue that the responsibility inflation problem faced by anti-luckists, could plausibly also apply to compatibilists. I argue that unlike the compatibilist, the modest incompatibilist position can deal with the inflation problem in an intuitive and principled way.

8.1 Fischer's Moore Shift

Given what I have argued so far, both modest incompatibilists of a Kanean persuasion and compatibilists have to acknowledge that everything we do is ultimately a matter of luck.[113] On both views, the only type of control that one has is compatibilist control, and compatibilist control cannot protect against ultimate luck. Few, however, have directly addressed the problem of ultimate luck. Fischer – a compatibilist – is one of the exceptions. He comes out fighting.

[113] And this would also apply to Valerian incompatibilist positions too, if they were still in the running.

Fischer does not try to deny ultimate luck.[114] We saw in Chapter 3 that Fischer is prominent amongst those contemporary compatibilists who use Frankfurt-style cases to deny that path-picking control is needed for responsibility. Instead Fischer highlights that "[w]e are at every point thoroughly subject to factors entirely outside our control ... [w]e are not even in a tiny bubble of control, but we are, in a sense, swimming in a vast ocean of chance and luck" (forthcoming *b*). He elucidates:

> Suppose my parents had beaten me mercilessly when I was very young, so that I had significant physical (neurological) and emotional damage. If the damage had been sufficiently bad, I would never have developed into an agent at all. And yet it is quite clear that I never had any control over whether my parents beat me in this way. Similarly for an infinitely large number of factors. For example, I had no control over whether I was born with a significant brain lesion that would impair or expunge my agency. Had I been born with such a lesion, I would never have developed into an agent at all, or would have developed into an agent with a very different character and set of dispositions. Again: I had no control over the fact that I was not dropped on my head (accidentally or deliberately) by my parents when I was very young. But had I been dropped on my head in a certain way, I would not have developed into an agent at all, or might have developed into a very different sort of agent.
>
> When one begins to think about this sort of thing, one quickly realises that we are incredibly lucky to be as we are. I had no control over the fact that I was not hit by a bolt of lightening when I was young (or, for that matter, yesterday), or that I was not hit by a meteorite, and so forth. But had any of these things occurred, I would not be the way I am today – and I certainly would not be typing this paper at my computer! Life is extraordinarily fragile, and (from a certain perspective) we are remarkably lucky to be agents at all, or the particular agents we are (with the particular dispositions, values, and psychological propensities we actually have). Intuitively speaking, I am not

[114] Dennett, another compatibilist who does discuss the issue of luck, takes a different, and altogether unsatisfactory approach. Dennett argues that to talk of ultimate luck is to misuse the term 'luck' – in this way his response is similar to some of Kane's in Chapter 5. But no reason whatsoever is given why we cannot talk of ultimate luck – we clearly *can,* the idea does not seem absurd at all, or in any way unintelligible (see Dennett 1984, pp. 92-100, and for discussion see Smilansky 2000, pp. 45-46 and Russell 2002b, sec. 5)

> "ultimately responsible" for my particular psychological traits or even for my very agency. We are not "ultimately responsible" for the "way we are", and yet it just seems crazy to suppose that we are thereby relieved of moral responsibility for our behaviour. (Fischer forthcoming *b*)

Concluding that we are not responsible is what some would, and have, concluded.[115] But the point Fischer makes is that it would be an equally legitimate dialectical move to take the widespread influence of luck to imply that luck-eradicating control cannot be what our common sense conception of responsibility presupposes.

> One could say that such a picture [in which we have luck eradicating control] is endorsed by commonsense but utterly impossible to fulfil. Or one could say that such a picture, being obviously and straightforwardly impossible to fulfil, cannot be the picture endorsed, upon reflection, by commonsense. The latter possibility... seems to me to be the path recommended by a certain sort of philosophical maturity and wisdom. Be that as it may, my more minimal point (to which I'd retreat if pressed) is simply that the latter approach is no less plausible than the former. (Fischer, forthcoming *b*)

The problem of ultimate luck, as has often been pointed out, bears resemblance to the problem of scepticism in general, where "conditions which seem perfectly natural, and which grow out of the ordinary procedures for challenging and defending claims to knowledge threaten to undermine all such claims if consistently applied" (Nagel 1979, p. 27 and see also Duff 1996, p. 332). What Fischer is suggesting above is a dialectical move analogous to Moore's shift regarding the problem of scepticism about knowledge. Rather than directly refute the sceptical challenge to knowledge, Moore switched to our common sense belief that we know we have a hand before us, and then asked us to recognise that by our everyday standards the sceptic's standards are unduly demanding (see Unger 1984 & 1986; Moore 1962). Granted, we do not have to take up the common sense everyday perspective and can instead view things from the more demanding

---

[115] Prominent amongst those who conclude and affirm that we are not responsible are Smilansky (2000; 2002) G. Strawson (1986; 2002; 2003) and Pereboom (1995, 2002). But there are also the control incompatibilists who believe that whilst we would be non-responsible if everything we did turned out to be a matter of luck, in fact we *do* have luck-eradicating control and so we are responsible after all (so these are control-libertarians). Amongst these we can number O'Connor (2002, 2003) and Clarke (2002,

sceptic's perspective. But the point is that one perspective is not the right one, and there's as much justification to take it as a fixed point that one knows one has a hand before one, as there is to take as a fixed point the sceptics demanding standard. What Fischer is doing is saying exactly the same thing regarding control and responsibility. We could see luck's involvement in all that we do as a problem, one which threatens to undermine responsibility. But then again, one can instead decide to see the ineliminability of luck as a sign that luck is not a problem.[116] As Smilansky puts it:

> While compatibilists like Dennett have tried to convince us that luck is not meaningfully present in pertinent cases, Fischer takes the opposite approach, in the attempt to neutralise the sting of luck. Since sceptics must agree, then their own luck-based argument is put in jeopardy. If we do not mind the necessary presence of luck in our lives, then why should we be worried about the threat of 'ultimate luck'? (2004, ms. p. 2 and quoted in Fischer *forthcoming* b. p. 9).

I believe that Fischer's approach has merit. Fischer's interest in making this argument is obviously to make us more hospitable to compatibilism. But in what follows I will be using essentially the same argument to call into question why incompatibilist control of the luck-eradicating variety is thought by so many to be necessary for responsibility. The standard rationale is that incompatibilist luck-eradicating control is needed precisely because of its luck-eradicating properties – an assumption that I term anti-luckist. But I think that considerations similar to those that Fischer has brought to our attention above render this rationale implausible.

2003), though Clarke seems currently to be in the process of giving up his libertarianism in favour of hard Incompatibilism.

[116] Nagel rules out this kind of approach in his famous article on Moral Luck saying that "[i]t would therefore be a mistake to argue from the unacceptability of the conclusions to the need for a different account of the conditions of moral responsibility. The view that moral luck is paradoxical is not a *mistake*, ethical or logical, but a perception of one of the ways in which the intuitively acceptable conditions of moral judgement threaten to undermine it all" (1979, p. 27). But I think Nagel is too hasty here. As I will point out in the next chapter, our intuitions about a number of cases – cases where it is tempting to invoke luck – can be explained in non-luck terms. So, for instance, 'resultant' luck intuitions can be explained in terms of our feeling that the appropriate ground for responsibility is the agent's character rather than the actual impact they make upon the world.

8.2 Restricted Control

Although it is hard to see how we could have incompatibilist control, it is uncontroversial that if we somehow could have genuine alternative possibilities and causal-luck eradicating path-picking control over which possibility is actualised, then we would have a solution to the problem of responsibility grounding free will (van Inwagen 2002, p. 168). Compatibilists, for instance, do not think that possession of incompatibilist control would undermine responsibility, they just do not think it is necessary for responsibility.[117] Most sceptics only believe it is impossible that conditions could obtain which would legitimise judgements of moral responsibility. But they would allow that if, per impossible, a certain combination of conditions *did* obtain, then we would be morally responsible. If we are focussing just on the control requirements for responsibility then it is uncontroversial that incompatibilist control is sufficient for responsibility. The free will debate is largely over whether such incompatibilist control is possible or necessary not over its sufficiency.

Bearing this in mind, and without taking back anything I argued in Chapter 5, I now want to draw attention to a mistake that many critics of incompatibilist control make. The mistake involves identifying incompatibilist control with something that it is not. To see this, it will first be useful to make use of a distinction drawn by Michael Zimmerman between restricted control and unrestricted control:

---

[117] It is true that some compatibilists – following Hobbes's lead - might pretend not to understand incompatibilist control in any sense. But I think this is disingenuous. I agree that incompatibilist control is mysterious, that does not mean that incompatibilist control itself makes no sense . We do, I think, known what the control incompatibilist is talking about. The following remarks by Honderich are, in my view, exactly right:

> Suppose I have no idea of why the petunias on the balcony *need sun,* but am persuaded they do, no doubt by good evidence. Despite the evidence, I have no acquaintance at all with photosynthesis, not even any boy's own science of the matter. It does not follow, presumably, that I lack the idea that the petunias *need sun.* I could have the idea, too, in a pre-scientific society where news of the science of the thing would for a long time make no sense. Could I not also have the idea, in a later society, if al of many attempts to explicate the need had broken down in obscurity and indeed contradiction?...
>
> The friends [of incompatibilist control] speak no nonsense when it transpires that they cannot in some way explain how it comes about that there is [incompatibilist control], or would come about if there were any. They still speak no nonsense in what went before if their attempts to explain are themselves pieces of nonsense. (2002, p. 474)

I assume in what follows that most compatibilists would accept that the incompatibilist is not talking nonsense. Rather, they simply dispute that incompatibilist control is actually possible, and whether it is actually needed.

> [W]e should distinguish two ways in which something may be beyond someone's control. Roughly, one may be said to enjoy *restricted* control with respect to some event just in case one can bring about its occurrence and can also prevent its occurrence. One may be said to enjoy *unrestricted* or *complete* control with respect to some event just in case one enjoys or enjoyed restricted control with respect both to it and to all those events on which its occurrence is contingent. Thus an vent may be beyond someone's control either in the sense that it is not in his unrestricted control or in the stronger sense that it is not even in his restricted control. (1993, p. 219 one note omitted)

Unrestricted control, as Strawson, Smilansky, Wolf, and many others point out, seems to be what one would ultimately need to satisfy the requirement for incompatibilist control (Strawson 2003; Smilansky 2000; Wolf 1990). I argued one can be led to this conclusion in Chapter 5. We first assume that essential to incompatibilist control is that it eradicates causal luck of the kind witnessed in roll-back cases. Then we argue that eradicating this luck requires a deeper self who is exercising this control, but that the problem of causal luck at this deeper level simply re-arises, and hence the regress begins. It would seem, following this reasoning, that to eradicate causal luck, the agent needs is control over the way that they are, and that this would require control over all those factors that contributed to the agent's being the way that they are. In other words, it looks as if we need unrestricted control. We do not have unrestricted control: it would involve having control over whether or not the sun rises, or whether or not we have a fatal stroke, and, more basically, whether or not we were born. Unrestricted control is plainly incoherent, "a sort of rape and perversion of logic" as Nietzsche put it (1966, p. 21). For to have unrestricted control – so control over *everything* that is in any way necessary or sufficient for one acting as one does – is to need control over coming into existence. This is impossible.

However, what we nevertheless need to recognise is that control-control, if we could have it, would not amount to having control over all those factors that contributed to an agent's being the way that they are. Unrestricted control is not incompatibilist control. No sane control-incompatibilist thinks that they actually have unrestricted control or

need it for moral responsibility.[118] For instance, one of the most vociferous critics of incompatibilism (and compatibilism), G. Strawson, accepts that very often we *do* take ourselves to have incompatibilist control of the sort needed for moral responsibility and of the sort that *he* thinks is impossible, and he gives the following example as an illustration:

> Suppose you set off for a shop on the evening of a national holiday, intending to buy a cake with your last ten pound note. On the steps of the shop someone is shaking an Oxfam tin. You stop, and it seems completely clear to you that it is entirely up to you what you do next. That is, it seems to you that you are truly, radically free to choose, in such a way that you will be ultimately morally responsible for whatever you do choose... as one stands there, one's freedom and true moral responsibility seem obvious and absolute to one. (2003, pp. 216-217)

In these cases we do not think that we have unrestricted control. In such cases we do not think that it is entirely up to us whether the sun rises or whether we continue to live. It might be objected that what I've said above misses the point in that unrestricted control is what is needed in order to satisfy the control incompatibilist's demands. There is no way of having path-picking control unless one has unrestricted control. This would be another way of saying that it is impossible to eradicate causal luck (or causal luck internal to the agent) without also having to eradicate all other sources of luck (including circumstantial and constitutive). The point is not that the control-incompatibilist actually believes that we have unrestricted control, or that the belief that we possess incompatibilist control is the belief that we have unrestricted control. Rather,

---

[118] Some have maintained that free will involves something like unrestricted control. Both Descartes and Sartre seem to, holding that our free will is absolute and unlimited by circumstances. Descartes, for instance, describes our free will as "so free in its nature that it cannot be constrained" (1976, p. 21). More recently Rogers Albritton has voiced his own sympathy to this kind of view. But I am at a loss to understand exactly what this means, and Albritton himself accepts that it is hard to make sense of. We certainly cannot, surely, be held to have control over all of our circumstances – but none of the above deny that. Instead what is being maintained is that our circumstances in no way constrain our wills. So even the above do not accord us unrestricted control in the sense that I am talking about above. It is not even clear that 'restricted' incompatibilist control as I understand it, does not also amount to the unrestricted control of Albritton and others. For in a sense incompatibilist control *is* unrestricted within certain boundaries – which sounds contradictory but isn't. However, I leave this matter unresolved – I simply dismiss the above views as being incomprehensible and I assume that most modern control-incompatibilists agree.

it is that when we try to analyse what this control has to involve we are driven to having to insist that we need unrestricted control.

But unrestricted control is not a solution to the control incompatibilist's problem. When we arrive at the demand for unrestricted control we do not think 'ah, yes, *that* is what incompatibilist control is". It is a mistake to identify incompatibilist-control with unrestricted control.[119] It is a mistake that makes the demand for incompatibilist control look unduly demanding. It gives the impression that incompatibilist control is a demand for a mammoth amount of control – when in fact it isn't at all. It is, however, conceptually demanding. But it is not demanding in requiring that we have a considerable amount of control.

The point is that control-incompatibilists and compatibilists do not differ very greatly in the degree of control they take themselves to have. Control-incompatibilists do not see themselves as something akin to Gods capable of raising and lowering the sun. The difference is not so much the degree of control that each takes themselves to have, but the kind of control. The kind the incompatibilists take themselves to have clears a small luck-free foothold, but that is all. The difference comes down to this for the compatibilists everything that one does is ultimately a matter of luck, whereas for the control-incompatibilist everything that one does is largely a matter of luck. For once we recognise incompatibilist control for what it is – a restricted form of control – we have to recognise that possession of such control still leaves us exposed to a large measure of circumstantial and constitutive luck. Everything that we do would be infected with circumstantial and constitutive luck. It is just that these luck factors would not be the whole story or explanation of why one did what one did.

8.3 Luck and Responsibility Inflation

I made two points above. Firstly, it is uncontroversial that if we could have incompatibilist control, and so if we could have a kind of control that eradicates agent

[119] Note it is not specifically the restricted nature of this control that makes it mysterious or incoherent. For if one takes away the restricted requirement, then one *still* has a demand for something mysterious or incoherent. For how could we possibly have an *unrestricted* power of self-creation? In other words, the 'restricted' nature of incompatibilist control does not make for a more incoherent mix than a plain requirement for unrestricted control. In the same way that the demand for a square circle is incoherent, the demand for a square triangular circle is not *more* incoherent.

internal causal luck, then we would satisfy control requirements for responsibility. Secondly, incompatibilist control is not a requirement for unrestricted control. Incompatibilist control is actually a very modest kind of control, for all its mystery. Possession of such control would clear a small luck-free foothold, a luck-free centre to our actions, but all that we do would remain heavily luck infected. In terms of exposure to luck, the difference between having just plain compatibilist control and having incompatibilist control as well is the difference between luck going all the way through, and most of the way through.

To get clearer about luck's continuing involvement, consider that path-picking control, if we had it, would not give us control over the nature or frequency of junctions. In other words, even if we could have incompatibilist control, this would not give us control over the option ranges we pick from, or how often we get to pick. The nature of our option ranges will be determined by our circumstances, and by our constitution. One cannot, for example, choose to do something that it did not occur to one to do. One cannot choose to do something that one has no desire or motivation to do. Yet what it occurs to us to do, and what we are motivated to do, are matters that would not be under our incompatibilist control (if we had it).

One can, to some extent, take control of these matters, by cultivating certain habits and so forth. But the extent to which one has the opportunity to do this will be a function of matters outside one's control. (As we saw Aristotle note in the previous chapter, whether or not one's efforts to cultivate certain habits are successful is again, a matter that is not under one's incompatibilist control.) One's initial desires and motivations, and what initially occurs to one – these are matters over which one cannot have exercised any incompatibilist control. As O'Connor – a control-incompatibilist of the agent-causal type – acknowledges:

> We enter the world with powerful and deep behavioural and attitudinal dispositions. Long before we mature to the point of making sophisticated, reflective choices, we are placed in environments that mould and add to those dispositions. Such factors heavily influence our early choices, even if they do not causally determine all of them. They certainly do determine that Billy will choose from only a very limited range of options in any given situation, a range

> that will differ quite a bit from that open to Susie under similar circumstances. These choices and continuing contingencies of circumstance, in turn, will sharply circumscribe the options Billy considers at a more reflective stage, when we begin to hold Billy accountable for his actions...
>
> Surely one must concede in response that responsibility for "shaping who I am" and for the choices that ensure from this comes in degrees and, indeed, can only sensibly be measured within a limited scope of possibilities. We cannot hold Billy responsible for failing to consider an option entirely outside the range of his experience. And his responsibility for passing by options that are within the range of his experience but that he has had precious little opportunity to consider as attractive is attenuated... Perhaps the important point to emphasise here is that the [control incompatibilist] can... accept perfect responsibility for one's choices and character is not just contingently lacking in human beings but is impossible: it would require perhaps perfect indifference at the outset, or at least an openness to all possible courses of action. The coherence of *that* idea is doubtful. (2002, pp. 351-352)

Our option ranges are determined by the interplay between our given character and our circumstances, and these are ultimately out of our control – even our incompatibilist control, if we had it. It is only within these boundaries that incompatibilist control can get to work. Even then, there is luck. For to exercise incompatibilist control requires that one stays alive or conscious. Again, this is not a matter over which incompatibilist control gives one any control. The platform from which we could operate incompatibilist control is floating in an ocean of luck. Even though our compatibilist control can be said to be enhanced by possession incompatibilist control, on reflection the enhancement in question is surprisingly slight. Everything that we do remains heavily luck infected.

I hold that recognition of the above calls into question the *rationale* behind the need for incompatibilist control. For it becomes implausible, once one recognises just how great luck's involvement is in everything we do, for incompatibilist control to be needed in its luck eradicating capacity: it performs so poorly on that score.

8.4 Response

The control incompatibilist might respond to what has been said so far by pointing out the difference between everything that one does being completely a matter of luck, and everything that one does being partially lucky is a very significant one. Incompatibilist control does secure a luck free foothold, and that this is all that is needed for responsibility. For in securing a luck-free foothold, incompatibilist control in effect renders all these other sources of luck *benign.*

This is unconvincing. Or at least it is unconvincing if one holds that the reason we need incompatibilist control is precisely to clear a luck-free foothold. For if one gives that kind of rationale – the anti-luckist rationale - then one cannot just ignore the manifold ways in which luck remains a huge factor in all that we do. If we have libertarian control then what we do is not completely a matter of luck. But as already pointed out (and acknowledged by O'Connor in the above quote) everything we do is also a (high) degree lucky. And if 100% luck rules out responsibility, it would be odd if 95% luck were no problem at all. For what kind of concern over 100% luck would not apply also to 95% luck? For instance, if we judge it unfair to blame someone for something that was 100% lucky for them, then it is also unfair to blame someone for something that was 95% lucky. Anti-luckism, in other words, commits one to taking luck seriously, wherever it impacts upon an agent's responsibility. If one refuses to do this – if one just insists that remaining sources of luck are benign and do not raise any issues – then one can hardly criticise the compatibilist for ignoring ultimate luck. For the only difference between oneself and the compatibilist will be that the compatibilist turns a blind eye to *slightly* more luck than you do.

In summary, if one holds that incompatibilist control makes us responsible because it eradicates some luck, then one cannot justify ignoring all the other ways in which luck is involved in what we do and are responsible for. But what would taking such luck seriously involve?

8.5 The Implications of Consistent Anti-Luckism

Michael Zimmerman has recently highlighted just what consistent anti-luckism yields (2002). As we saw above, Zimmerman usefully distinguishes between restricted and unrestricted control. He agrees that responsibility presupposes only some form of restricted control. He is ambiguous over whether that control need be incompatibilist or compatibilist, but for our purposes we can assume that it is incompatibilist (to the best of my knowledge, Zimmerman *is* an incompatibilist, though for reasons to do with luck, he thinks it does not really matter which side one takes). Zimmerman is also anti-luckist and believes (quite consistently) that to the extent that luck factors play a role in what we do, such influence needs to be 'neutralised'.

He argues that because responsibility only plausibly presupposes some kind of restricted control, then the widespread involvement of luck in all that we do does not imply that our responsibility is subverted for what we actually do. In other words, luck's involvement does not prevent responsibility getting off the ground, for luck's involvement is consistent with our having some species of restricted control, and restricted control is all that it is plausible to suppose responsibility requires. But nevertheless, luck does need neutralising. Luck's involvement is not, in other words, benign. Its involvement renders our practices radically unfair. The general principle that Zimmerman invokes is as follows:

> [I]f a) someone's being *F* (where '*F*' designates some complex property comprising both epistemic and metaphysical components) is sufficient for that person's being morally responsible to some degree x, then, if b) it is true of S at some time that he or she would be *F* if *p* were true, and c) *p*'s being true is not in S's control at that time, then d) S is morally responsible to degree x. (2002, p. 565 fn. 53)

The implication of this principle is that we are morally responsible not just for what we freely do in the actual world, but for all that we *would* have freely done had factors not under our control been different. Thus, Jones the assassin does not get off the hook simply because a strange thought occurred to him, diverting him from deciding to pull the trigger. If Jones would have made that decision had the luck factor (the strange

thought occurring to him) not occurred, then he is as morally responsible *as if* he had made it, even though in the actual sequence he did not. If Jones in the actual world, succeeds in assassinating the prime minister, then he is responsible for having done this, but so too is his counterpart Jones* in a nearby possible world in which a gust of wind diverted the bullet.

Furthermore, an agent is morally responsible not just for the decision that they make in reaction to the circumstances in which they actually find themselves, but also those decisions they would have made had they faced different option ranges. With regard to constitutive moral luck the agent is morally responsible not just for the decision they make, but also for all those decisions they would have made had their background character traits been different. In other words, the implications of luck neutralisation are radical in the extreme. It "opens up the floodgates, as it were, when it comes to ascriptions of responsibility – of laudability as well as culpability" (Zimmerman 2002, p. 370). Your responsibility ceases to be restricted to what you actually do, but expands to include all that you would have done. What you actually do is just one small piece in a much larger picture. Your morally upstanding acts in the actual world can be seen as just one vote in a massive election in which *all* of your possible selves are voting.

It might be objected that in the cases where the agent fails to make a particular decision due to factors not under their control, there is nothing for which they can be morally responsible. In the case where Jones is unconscious at t2 it makes no sense to talk of Jones being morally responsible for what he would have done. What exactly is Jones being held morally responsible for? By hypothesis he did not actually do anything. In answer, Zimmerman proposes we should distinguish between the scope and degree of someone's moral responsibility (2002, p. 560). Jones, we can say, is responsible for less than he would have been had he not been rendered unconscious. But that is a matter of the scope of his moral responsibility, not its degree. Although he is morally responsible for less, he is morally responsible to the same degree as he would have been had he not been rendered unconscious. The moral luck that requires neutralisation is that which concerns the degree of an agent's blameworthiness, not its scope. The objection was that there is nothing that Jones could be morally responsible for, given that in the actual sequence he does nothing at all at t2. But now we can say that whilst the scope of Jones' moral responsibility has "dwindled to nothing" his degree of moral responsibility

remains the same (Zimmerman 2002,p. 564). Jones, we can say "is responsible; he is just not responsible for anything ..[h]e is, as I shall put it, 'responsible tout court'" (Zimmerman 2002, p. 564). He explains:

> Lest this appear unduly paradoxical, let me hasten to add that it is nonetheless the case that [Jones] is responsible in virtue of something, and this something just is his being such that he would have freely [decided to X], had he had the cooperation of certain features of the case. All responsibility, including responsibility tout court, is fundamentally relational. It is precisely because [Jones] is responsible in virtue of the very same sort of fact (the fact that he would have freely [decided to X], had he had the cooperation... of certain features of the case) in virtue of which [Jones* - counterfactual Jones who does successfully make decision X] is responsible, that [Jones and Jones*] are responsible to the same degree. (Zimmerman 2002, pp. 564-565)

Note, if one were a compatibilist then it looks as if the consistent application of Zimmerman's principle would seem to threaten to render everyone's responsibility *tout court* the same. In other words, everyone would be as responsible as everyone else. For it would be axiomatic that Jones would have done what Susie did if he had been in her exact shoes (where this involves having her character and history in the exact circumstances in which Susie found herself), and it was not a matter under Jones's control that he was not in Susie's shoes (Smilansky 2000, p. 45).[120] In other words, responsibility would be dissolved (see Zimmerman 2002, pp. 570-571).[121]

---

[120] If one is a 'free will either way' compatibilist and one sometimes faced torn indeterministic decisions, then Zimmerman's principle would imply that one is responsible not just for the decision that one actually made, but also the decision that one could have made were it not for one's causal luck.

[121] From a compatibilist's perspective it is an interesting question whether the implications of Zimmerman's view are such as to render the compatibilist in the same position as the hard determinist. Is everyone being as responsible as everyone else equivalent to no-one being responsible for anything? I think that the answer here is no, for the following reason. Assuming that in practice we would continue to punish and reward irrespective of whether hard determinism or compatibilism is true (most hard determinists still think it *justifiable* to punish people – it is just that the justification is normally Consequentialist rather than desert-based), then in the former case punishment would at still respond to an agent's desert, whereas in the latter it would not. It would not respond to a person's full desert – for that to be the case we would all have to be punished and rewarded equally, which is clearly practically impossible. Nevertheless, the punishments that people receive will still be deserved, even if it is the case that those who receive such punishment deserve *more* punishment than they receive. Getting part of what you deserve is still getting something you deserve. But note that the view could be practically equivalent to a consequentialism or hard determinism, for there would be no reason to restrict blame and praise to what the agent actually does. However, the scapegoating objection to straightforward consequentialism would not apply – for we would not be innocent of those matters for which we are being punished.

Zimmerman does, however, acknowledge that there are limits to neutralisation – limits placed by what one considers *essential* to the agent. There is something incoherent about saying what Jones would have done if he had been a different person (Zimmerman 2002, pp. 574-575; see also Statman 1993, pp. 12 - 13; Rescher 1993, pp. 156-157). This means that "the role that luck plays in the determination of moral responsibility may not be entirely eliminable... because, regardless of just which personal characteristics should be said to be essential to persons, it is presumably correct to say that some are" (Zimmerman 2002, p. 575). But as Zimmerman acknowledges, whilst this may show how some limits can be put on neutralisation, it remains the case that if one is anti-luckist one should still neutralise to the greatest extent possible, within the boundaries set by what is essential to the agent.

Another objection is that neutralisation presupposes that there is some fact of the matter about what an agent with incompatibilist control would have done. For if we can say for certain what the agent would have decided, then we must presuppose a necessary connection between the agent's antecedent circumstances (their motivational states, desires etc.) and their subsequent decision. This is fine if one is talking about compatibilist free will. However, this necessity will be missing if the agent has incompatibilist free will. In other words, one cannot say what the agent *would* have freely decided if free will is understood in incompatibilist terms, precisely because what they would have decided is not determined by prior circumstances (Zimmerman 2002, pp. 572-573). There is no truth about what an agent with incompatibilist control would have done in different circumstances.

It is debatable whether there can be true counterfactuals when it comes to agent's incompatibilist free acts – Plantinga thinks there can be, whilst others think not (Plantinga 1974, p. 173; those who do not: Adams 1977; van Inwagen 1997; Hoffman and Rosenkrantz 2002, ch. 6). This is a difficult and controversial matter, but I assume that those who would want to block luck neutralisation would not mind being committed to denying the truth of such counterfactuals. So it seems that there *is* a way for the anti-luckist incompatibilist to block neutralisation.

But this is a rather small victory. Even if there cannot be true counterfactuals about incompatibilist free acts, this is no real help when it comes to addressing the underlying unfairness to which neutralisation was the proposed answer. For denying that there is any truth value to relevant counterfactuals does not magically mean that agents are *not* getting off the hook by luck or that our practices are fair. It leaves us perplexed about how one could adjust for such luck. It means we just have to live with a very large measure of unfairness.

The above point about there being no fact of the matter applies only to directly free acts. But as was noted in the previous chapter, unless one is going to take up a restrictivist position, most incompatibilists will allow that an agent can be responsible for their indirectly free acts just so long as they trace, in appropriate ways, to an earlier directly free act. Neutralisation would apply to indirectly free acts. For example, Jake freely decides to get drunk and exercises incompatibilist control over this decision. Whilst drunk, he gets into a fight and breaks someone's leg. Jake's act of breaking someone's leg was *not* a directly free act, for Jake, we can assume, did not exercise incompatibilist control over his doing this. But the tracing view allows one to say that Jake is responsible both for getting drunk, and for breaking someone's leg because the latter act traces to Jake's directly free act of deciding to get drunk.

Yet it is only thanks to Jake's circumstantial luck that he broke someone's leg, rather than killed someone. Because we are dealing with an indirectly free act there is no problem in saying that Jake would have killed someone in different circumstances. This is because only Jake's directly free act was indeterministic. Jake's indirectly free act was (or could have been) necessitated by antecedent circumstances. So, tracing views, which save the incompatibilist from having to limit responsibility to directly free acts, now exposes the anti-luckist incompatibilist to neutralisation.

The anti-luckist incompatibilist could restrict responsibility to directly free acts, and in this way avoid the neutralisation problem. But this would be to seriously limit responsibility. Furthermore, as I pointed out above, this would not do anything to demonstrate that the agent's directly free acts are not heavily luck infected. Thus the anti-luckist would just have to accept a very large measure of unfairness by their own

lights. The only selling point of this view would be that matters are *slightly* less unfair than with compatibilism. But that is hardly a ringing endorsement for the position.

In short, once one takes anti-luckism seriously, neither compatibilism or incompatibilism look attractive, precisely because of the relative ineffectiveness of incompatibilist control at eradicating luck. If we could have incompatibilist control, we really would not have much more control than if we have just compatibilist control.[122]

8.6 An Alternative Rationale for Incompatibilist Control

The upshot of my discussion above is that if one is going to take the anti-luckist rationale for incompatibilist control seriously, one is going to find oneself driven to making the kind of radical revisions to our common sense conception of responsibility outlined above. Alternatively one is going to have to seriously restrict responsibility and at the same time admit that incompatibilist control works only to make matters slightly fairer than they would be if we only had compatibilist control. Such considerations cast serious doubt on the plausibility of the anti-luckist rationale for incompatibilist control.

Here then, we can take Fischer's lesson. We can see luck's deep involvement in all that we do as indicating not a problem, but rather luck's non-toxicity. In other words, we make the Moore-Shift. But how might we have come to see luck as toxic to responsibility?

In the previous chapter I stressed the value of being, to use Kane's phrase, an independent source of activity and of thereby having a degree of ownership over what one does and becomes. I also argued that agent-internal indeterminism, unsupplemented by incompatibilist control, would secure these goods. But I noted in Chapter 6 that there

---

[122] Consider also that it is now widely accepted that you cannot tell through introspection whether you have incompatibilist control (see Double 2002; Kane 2005; Chisholm 1976; Taylor 1966; van Inwagen 1983). Our internal experience of controlling our decisions is consistent with determinism being true. There's a case for saying that our internal experience would be *illusory* if determinism were true – for we sometimes (or some of us anyway) take it that we have genuine alternative possibilities. But in the first place, that seems to vindicate modest Incompatibilism of the Kanean sort rather than control-incompatibilism, and furthermore this does not demonstrate that we actually *have* incompatibilist free will. In short, coming into possession of incompatibilist control is not something that one would be aware of – it is not as if one would suddenly feel 'empowered'. And this just testifies to the modesty of this control. We can have it, without realising, and we can lose it without realising. Given just how important some consider possession of this control to be, it is rather odd that it can be so indiscernible.

is a legitimate, though mistaken, concern that agent-internal indeterminism would erode our control. This suggests the possibility of giving the following story.

We want the genuine alternative possibilities that agent internal indeterminism offers in order to secure independence and ownership. But then we reflect that the genuineness of the alternative possibilities threatens our compatibilist control. It is for *that* reason that we think a new kind of control is needed. Incompatibilist control is something we take ourselves to need in an open-future world in order to protect against control-erosion, but it is not the reason we want to be in an open-future world. It was securing those other goods – the goods of independence and ownership – that we were after. After all, what value is there in having incompatibilist control as opposed to just compatibilist control if it is not to do with being independent and having ownership? On this story the rationale behind incompatibilist control is that it is needed in a restorative capacity. Compatibilist control is *not* inadequate for responsibility *per se*. It is just that we do not think that we can have compatibilist control and genuine alternative possibilities at the same time, and hence the need for incompatibilist control.

The story I have suggested above strikes me as eminently plausible. Yet it is overlooked. Nearly all control-incompatibilists see genuine alternative possibilities as primarily just clearing room to accommodate incompatibilist control. So their idea is not that incompatibilist control is restorative. Rather it is that possession of path-picking control yields an overall enhancement of control in some absolute sense – an enhancement over and above what could be achieved in deterministic settings.[123]

Similarly with the compatibilist. When the compatibilists such as Fischer and Haji point out that alternative possibilities by themselves would not enhance control they presuppose that enhancing control could be the only reason alternative possibilities would be part of the necessary requirements for responsibility. In other words, they

[123] For instance, Randolph Clarke is a control incompatibilist who nevertheless agrees that agent-internal indeterminism of the kind that one finds with Kane's view, does not erode compatibilist control. He would, in other words, agree with the conclusions I drew in Chapter 5 and Chapter 6. However, when it comes to moral responsibility he thinks that whilst "[s]uch a view secures a type of leeway or openness not available in a deterministic world,... the view provides the agent with no additional positive power to determine what he does; it does not secure any greater degree of active control. And this is what seems to be needed I there is to be a different verdict concerning desert and hence responsibility" (2002, p. 376). In other words, Clarke thinks that compatibilists control needs supplementing if we are to have the kind of free will necessary for moral responsibility.

presuppose that the only problem anyone could have with compatibilism, is over the adequacy of compatibilist control.[124]

So, much of the contemporary debate over free will has remained a debate over whether or not compatibilist control is adequate for responsibility, or whether it needs supplementing. In other words, the debate has tended to be over how much control we need, and which side offers the most and at what costs. In the classic debate it is accepted that the incompatibilists offer more control than the compatibilists, but at heavy conceptual costs, whilst the compatibilists offer less control, but at least offer a more sober, 'down to earth' control.[125]

But according to the story I have suggested compatibilist control is adequate, it is just that free will is not only about control. We want to be the ultimate sources of our actions, and this is a basic demand, not one made in order to increase our control in any absolute sense. And there is an understandable (albeit mistaken) concern that the alternative possibilities needed to satisfy the ultimate source demand will erode our control. For that reason there is felt to be a need for a different kind of control – path-picking control.

I do not have any way of proving that this story is correct. But it does seem to me at worst every-bit as plausible as the more usual anti-luckist rationale. If we accept my story above we should come to see that a lack of incompatibilist control is a problem only if it is true that having genuine alternative possibilities does erode compatibilist control in some significant way. I have argued in Chapter 6 that it does not. So, whilst it is perfectly reasonable to think that agent-internal indeterminism (so, genuine alternative possibilities) might pose a risk to our control, there's a case for saying that such fears

---

[124] We see this very clearly in Fischer. Whilst acknowledging that even if Frankfurt-style cases work and PAP is refuted, there is still room for an incompatibilist to insist that we need to be the originators of our actions, Fischer says that he sees "no promising strategy for arguing that causal determinism threatens moral responsibility apart from its allegedly ruling out alternative possibilities" (1994, p. 151). He only sees alternative possibilities as plausibly required, if required to supply incompatibilist control (1999a).

[125] As Double says on behalf of the compatibilists (Double is not a compatibilist):

> We compatibilists are *magnanimous* in the face of determinism; we accept nature, recognise that we are part of it, and are at peace with it and ourselves. You libertarians [incompatibilists who affirm free will] – who want to carve out a special niche in the cosmos for free agents, however incomprehensible it might be – are insufficiently magnanimous. You are vain, antagonistic toward the cosmos, and too quick to fly to extravagances to aggrandize yourselves.

are unfounded. Similarly, there is a case for saying that incompatibilist control is not really needed after all. We can explain why incompatibilist control is associated with responsibility, and we can explain why it is not actually needed after all.

8.7 Frankfurt-Style Cases

What I have argued here chimes-in with what I argued in Chapter 3. For in Chapter 3 I argued that Frankfurt-style cases do not call into question the validity of PAP. They do not call into question the idea that for responsibility we require genuine alternative possibilities and independence from the past. However, I did allow that Fischer's variation of a Frankfurt-style case did cast doubt on the requirement for incompatibilist control. In other words, what Fischer's variation actually calls into question the reason alternative possibilities are associated with free will and responsibility. Unless one is already disposed towards compatibilism, then Fischer's Frankfurt-style case signposts the kind of position that I have been defending.

8.8 Compatibilism, Indispensability and Neutralisation

Fischer, whose basic approach to the question of luck I endorsed, is a compatibilist. As I pointed out at the outset of this chapter, the problem of luck is one shared by both compatibilists and modest incompatibilists, once we accept that compatibilist control is the only kind we have. But once one accepts the compatibility of luck and responsibility, why not just be a compatibilist? This is certainly where Fischer intends that his argument should lead us:

> [M]y suggestion is that, once one sees that the picture that favours [Unrestricted Control] is seen to be inflated and illusory, one might have less inclination to accept an incompatibilistic source requirement of *any* sort for *any* reason. That is, once one sees that there are a huge (presumably infinite) number of factors which are entirely out of my control (like the sun's coming out to shine) tat are such that, if they were not present, my agency would be very different or not even non-existent, one might be less inclined to object to

---

You should accept your freedom as it is and try to enhance it as you can, rather than engage in fanciful speculations. (1996, pp. 138-139)

> (or find problematic) the fact that, if causal determinism obtains, there will be a condition entirely "external" to the agent and over which he has not control which is causally sufficient for one's behaviour. (Fischer forthcoming *b*)

There are a number of reasons to stop short of compatibilism. Firstly, there are those problems to do with CNC control and ownership about which I have said enough already. But there is also another reason, this time it is to do with the very luck that Fischer thinks is not a problem. Let us return to Zimmerman's view that whilst luck does not preclude us having control over what we do of a kind sufficient for responsibility, it would nevertheless be unfair not to neutralise luck. I argued that if one is anti-luckist then one has to take this kind of unfairness seriously. However, there's a case for saying that we should take this kind of unfairness seriously even if we are not anti-luckist. Even if one is a compatibilist, and so even if one accepts ultimate luck is consistent with one having the requisite control over what one does, it surely remains the case that there is something unfair about one's moral responsibility being restricted to what one actually does, rather than all that one would do.

How might Fischer respond to the threat of luck neutralisation? Well, he could point out that neutralising luck in this way would yield a situation in which we are all virtually as blameworthy and praiseworthy as each other, such that responsibility would be in effect dissolved – a point made above. There is something to this response. But I believe that the modest incompatibilist has a better one. The modest incompatibilist can offer a principled reason why luck cannot be neutralised.

To see this, return to roll-back cases. In a roll back case we hold everything fixed prior to decision making, and then run the sequence through an innumerable number of times. If modest incompatibilist free will conditions are met, then in some re-runs the agent will decide one way, and in some another, at least on the occasion of torn decision making. If we apply Zimmerman's luck neutralising policy, then the agent's responsibility should not be restricted to making the decision that they made in the actual sequence. They should also, in fairness, be responsible for the decision that they would have made, had it not been for their differential causal luck (and also responsible for the character they would have developed had their causal luck been different). In

other words, when we are torn over which decision to make, we are responsible for both irrespective of which we end up making.

Note that if we do neutralise luck in this way, then the agent will lose the independence and ownership that the indeterminism was brought in to secure. For now the agent's moral responsibility will be determined entirely by factors external to their conscious will. Their moral responsibility will be a function of their circumstantial and constitutive luck. It was after all, not down to the agent that they were torn between deciding to X, or deciding to Y. All that it was ultimately down to the agent to do, was decide one way or the other. But the option range itself – that was not their doing.

So, with the modest incompatibilist view we have a principled reason why luck cannot be neutralised. If luck were neutralised one would lose the independence and ownership that are, in their turn, necessary requirements of being responsible at all. This brings out an important difference according to whether one puts the stress on control or ownership. If one's focus is control, then luck neutralisation looks unblockable. Granted, the compatibilist can always point to the fact that ultimate luck cannot, by hypothesis, be a problem. But this kind of response will only go so far. This kind of response will only help you resist the suggestion that luck prevents responsibility getting a foothold. But it will not really help one to argue against the unfairness charge. But if instead one puts the focus on ownership then we *can* block neutralisation, precisely because neutralisation would destroy the conception of ownership in question. This only holds for an incompatibilist conception of ownership – but then we have seen that there are independent reasons for favouring such an account.

My view has this advantage because it makes luck part of the solution rather than part of the problem. Luck is not a threat to responsibility, but part of why we are responsible. It is only through doing this, an admittedly very radical move, that neutralisation can be avoided. Compatibilists, either ignore ultimate luck (as in Dennett and Haji's case) or they argue that it is not really such a problem (as with Fischer). But only the modest incompatibilist has, if they've courage enough – courage that I have tried to muster here – the resources to be able to celebrate luck, and give it a positive, responsibility affirming role. For on my take on modest incompatibilism we are no longer morally

responsible despite our exposure to luck; we are responsible because of our exposure to luck. Luck of a certain sort, procures ownership for us.

To some what I have just said will sound preposterous. Kane and other modest incompatibilists at least have the decency to be embarrassed about the fact that their views leave one exposed to ultimate luck; Kane tries his best to hide the fact, as we saw in Chapter 5; van Inwagen hopes that there is something wrong with the luck objection and appeals to mystery in the mean time. I have openly admitted ultimate luck and made a virtue out of it, and in this respect I have gone even further than a compatibilist such as Fischer is willing to, and placed myself beyond the pale.

But in response, we can return to a point made in Chapter 5. Kane pointed out that when we talk of luck we usually mean "outside [our] control" (2003, p. 305). And it is plausible that we normally take luck to be toxic to responsibility precisely because we identify luck with being subject to external forces. In other words, to the extent that one's action was a matter of luck, it was a matter determined by factors external to oneself. If this is right, then luck *inside* the locus of our control should be treated with great care. We should not be too quick to assume that our intuitions about the toxicity of regular 'external' luck can be translated to 'internal' luck. For example, return again to the torn-decisions which we all face, and which Kane (and myself) believes to be the occasion of indeterminacy in our wills, and thus the occasion of the objectionable luck in question – the luck which I am saying does positive work, but which Kane and others try to paper over. When we make a torn decision, can we not readily admit that it was a matter of luck that we decided as we did? When I make a torn decision it seems quite plain to me that it was a matter of luck that I decided as I did. I admit that on such occasions I might just as easily have decided otherwise – that's just what it is to be torn about what to do. In fact, even if we have incompatibilist control over our decisions, surely our torn decisions are still lucky, insofar as we make them arbitrarily? Again, that is my experience of making torn decisions. Note, I am not saying that I experience a lack of incompatibilist control. I may, for all I know, exercise incompatibilist control over my torn decisions. Similarly, I am not asking that the conclusion of Chapter 6 be revoked. I do control my torn decisions in the compatibilist sense. But I would still say that I decided as I did by luck. It does not really matter what kind of control I have, my

torn decisions are matters of luck, because there's a very real sense in which I just arbitrarily decide. Yet at the same time, evident though the luck of my torn decision is, I would not for one instant think that I was not responsible for having made it. I would not offer, and nor would I accept as an excuse, that 'I could just as easily have decided otherwise'. For whilst my decision was arbitrary, it was still *my* decision, attributable to me and me alone. Matters would be different if it could be shown that me decision was not really down to me after all. If it could be shown that my decision was just the inevitable consequence of a process beyond the operation of my conscious will, *then* I would deny responsibility. But the mere fact that I could have arbitrarily decided otherwise *that* does nothing to suggest to me that I am not responsible. Quite the opposite.

8.9 Conclusion

My strategy in this chapter was to use a variation of Fischer's recent Moore-Shift argument to cast doubt on the rationale for incompatibilist control. I argued that whilst possession of incompatibilist control, if we could have it, would make us uncontroversially responsible, it would still leave us heavily exposed to luck. It would, it is true, eradicate some luck, but not much. This is a serious problem if one insists upon the anti-luckist rationale for incompatibilist control. For one will find oneself committed to grossly distorting our received conception of responsibility. In short, one will be impaled on the problem of responsibility inflation.

I suggested instead what I took to be an equally plausible story behind incompatibilist control. Incompatibilist control, I conjectured, might be thought needed only in some restorative capacity. Free will and moral responsibility, on this story, are actually associated with the satisfaction of a more fundamental demand than one of control: the demand for independence from the past, and so ownership over what one does and becomes. Incompatibilist control is brought in through fear that without some additional control, any gain in independence will incur a corresponding cost in one's control. I had already argued in Chapter 6 that the concerns over control erosion were misplaced, and as such those arguments should persuade at least some, that incompatibilist control might not be needed after all. Even if my arguments in Chapter 6 are not found persuasive, appreciating incompatibilist-control's new role may persuade some that it is

not needed anyway – for they might hold that the gain in independence is worth the cost in terms of control.

The important point though, at least where luck is concerned, is that once the focus is on ownership rather than control one is no longer imperilled by one's anti-luckism. One no longer faces the problem of responsibility inflation. One is better equipped than any other free-will theorist to explain in a principled way just why luck cannot be neutralised. Because on this view, luck is actually seen in a positive light, as something which is not always toxic but can actually provide us with something that we value: the independence to make a real difference through our choices, and to in this way see ourselves as significant in a way that can legitimise judgements of responsibility.

Chapter 9

Concluding Remarks

In these closing comments I will summarise the arguments of the preceding chapters and end with some remarks about the wider implications of some of the conclusions I have drawn.

9.1 Retrospective

In Chapter 1 I presented the problem of free will and framed it in terms of luck. If determinism is true then everything we do is ultimately a matter of luck, and so too if indeterminism is true. This seems to rule out free will of the kind needed for moral responsibility, because moral responsibility presupposes control.

In the following three chapters I focused on the threat from determinism. Our intuitive picture of free will involves having open pathways into the future and path-picking control over which way we go. Determinism rules out alternative possibilities and so seems to rule out possession of this kind of free will. In Chapter 2 I considered some standard compatibilist responses and I sketched an account of compatibilist 'reason responsive' control. In Chapter 3 I considered in detail a challenge to the intuitive picture of free will. The challenge is posed by Frankfurt-style cases. I argued that such cases fail to provide any reason to disassociate alternative possibilities and responsibility grounding free will. However, I did allow that Fischer's variation of a Frankfurt-style case did provide some, albeit rather flimsy, evidence that compatibilist control might be all the control needed for moral responsibility. But free will is about more than just having control, and so even if there's a case for saying that compatibilist control is the only kind needed, that does not in itself provide evidence against the intuitive idea of free will involving alternative possibilities. In Chapter 4 I went on to provide what I considered to be the major problem for compatibilism, and the primary threat from determinism. I followed others in arguing there is no relevant difference between cases in which an agent is subject to moral responsibility subverting clandestine manipulation and cases in which an agent's actions are causally determined. Because clandestine manipulation does not disrupt an agent's compatibilist control, the reason such

manipulation (and determinism) subverts responsibility must be for reasons to do with ownership. And it is reasons to do with ownership that account for at least part of the attraction of the open pathways conception of free will. For if we have genuine alternative possibilities then we have a degree of independence from the past and so can be said to be, to some extent, the ultimate sources or originators of what we do.

I then turned my attention to incompatibilist conceptions of free will. In Chapter 5 I argued that incompatibilist attempts to provide us with path-picking control of a kind that would protect against ultimate luck are either incoherent, or embarrassingly mysterious. Attempts to show how unassisted indeterminism can provide any enhancement to compatibilist control were shown to fail. However, in Chapter 6 I argued that unassisted indeterminism does not damage compatibilist control of the reason responsive variety. In short, indeterminism neither increases nor decreases our control. This means that in terms of exposure to ultimate luck, the incompatibilist is in the same position as the compatibilist. The only kind of control we have over what we do is compatibilist control of the reason responsive variety. But this does not put incompatibilism on a level with compatibilism. In Chapter 7 I argued that provided it is found in the right locations, indeterminism can provide the kind of independence and ownership that we saw determinism rule out in Chapter 4. This combination of compatibilist control with genuine alternative possibilities yields a modest incompatibilist position. What is interesting about this view is that it makes a certain kind of causal luck – agent internal causal luck – a source of free will rather than a threat to it. However, whilst this kind of modest incompatibilist position is more attractive than compatibilism, there remain persistent concerns about exposure to ultimate luck. Maybe luck can be a source of ownership, but many would still want to say that luck rules out moral responsibility. It was to these concerns that I turned in Chapter 8. There I argued that incompatibilist control – path-picking control – would do very little to reduce our exposure to luck. Yet the anti-luckism motivating the need for such control would commit us to neutralising luck. This would involve radical changes to our practices, rendering them to all intents and purposes unrecognisable. In particular holders of anti-luckist views are exposed to the problem of responsibility inflation. I suggested that it was not sensible to insist upon an anti-luckist rationale for path-picking control and that there was an alternative rational available. I also suggested that if path-picking control is needed, it is needed in some restorative capacity. But once this

alterative rationale is adopted, then one loses the motivation to reject straightforward modest incompatibilism. Just as modest incompatibilism is a more attractive view than compatibilism, it also appears more attractive – or at least *as* attractive – as 'control' incompatibilism.

9.2 Reservations

I have dismissed the need for path-picking control of the type associated with mysterious extra factors. Or rather, I have been seeing just how far we can go without them. But I do not pretend that my arguments on this front are decisive. I admit that the modest incompatibilist view that I advocate is more attractive when compared to compatibilism than it is when compared to control incompatibilism. In other words, I think a compatibilist has more reason than a control incompatibilist to move to modest incompatibilism (I myself moved from compatibilism to modest incompatibilism). The control incompatibilist's conviction that compatibilist control is inadequate for moral responsibility is unlikely to be given up easily. I have presented what I think are the strongest arguments showing the adequacy of modest incompatibilist free will. But I accept that in the final analysis we might still feel that indeterminism by itself cannot do the trick of making us truly morally responsible.

Nevertheless, if one does come to this conclusion my efforts will not have been wasted. I would still hold that my arguments for the priority of ownership are good ones. The same goes for my arguments for rejecting anti-luckism. The conviction that we still need something extra in terms of control does not call into question either of my claims above. In other words, I hold that even if we remain convinced that we need incompatibilist control, we should accept that our need for such control does not stem from the toxicity of luck. The committed control incompatibilist should see the need for incompatibilist control as basic. One either sees that it is needed, or one doesn't – nothing more can be said on the matter. The mistake that control incompatibilists tend to make is to offer a justification for the need for incompatibilist control: a justification that invariably makes mention of luck and thereby commits the control incompatibilist to anti-luckism. My arguments highlight how this is a mistake. These arguments hold good even if one does not see them as capable of showing modest incompatibilism to be an improvement on control incompatibilism.

9.3 Wider Implications

If what I have said in this thesis is right, then one upshot is that ultimate luck is not toxic to moral responsibility. What, though, are the practical consequences of this position? The modest incompatibilist position does involve recognising that our moral responsibility is a matter of degree. In this respect it echoes compatibilism. Though strictly speaking there is no reason why a control incompatibilist cannot accept that moral responsibility is a matter of degree also. In one sense everything is left as it is.

Consider resultant luck. It is a much debated matter whether our moral responsibility should extend to the consequences of our actions. For instance, should the reckless driver who accidentally kills someone be deemed more blameworthy than the reckless driver who, by luck, kills no-one? The view that I have defended in this thesis is agnostic on this question. And that, I hold, is just as well. The simple fact is that it is not clear what the answer to this question should be. If a view on free will comes down clearly on one side rather than the other, then this view paints itself into a controversial corner. Modest incompatibilism does not come down on either side. But it still offers guidance. For if ultimate luck is not a problem, indeed, if luck is actually required for free will, then we know not to try to settle this question using luck-based arguments. In other words, in trying to determine whether resultant luck should be allowed to affect blameworthiness, we should not make recourse to luck arguments. We should not argue that because it is just a matter of luck that one reckless driver killed someone whilst the other didn't, that *this* shows why we should blame both equally. To go down this road is to endorse anti-luckism, and this, as we saw in Chapter 8, is a serious mistake. (As I just mentioned, this is a serious mistake even if one remains convinced that we need incompatibilist control for moral responsibility).

Once we drop luck-based arguments the discussion over resultant luck can still continue. For one can argue that moral responsibility is responsibility for one's character. And as such there is no relevant difference between the two reckless drivers. Both show the same vicious character traits, and our tendency to react differently depending upon the consequences of an agent's actions reflects certain epistemic limitations. We do not have access to people's thought processes and intentions, and so

have to rely upon what a person's actions as a guide (Richards 1993; Rosebury 1995; Thomson 1993). But just as some might argue this way, others can argue that we are morally responsible for our actions where this includes their consequences. Still others might take a more Kantian line and argue that we are morally responsible for our mental acts – our acts of will – rather than our background characters or the consequences of our actions. The important point is that whichever way one goes here – whether one takes responsibility to be limited to character or whether one takes responsibility to be determined by what one actually does (where this might include the consequences of what one does) – it should be determined by matters independent of considerations of luck. That is the contribution the modest incompatibilist position can make to the wider debate over moral luck. It draws attention to the non-toxicity of luck, and in that way refocuses the debate over moral luck rather than settling it.

9.4 Conclusion

The problem of free will has been described as "excruciating" and "the most difficult problem in philosophy" (Double 2005, p. 24; Wolf 1990, p. vii respectively). Despite millennia of discussion it is often hard not to agree with Nagel when he says that "nothing believable has... been proposed by anyone in the extensive public discussion of the subject" (2003, p. 231). It is also hard not to agree with Wolf when she warns that "only fools rush in, at this point, thinking they have something to say about it" (1990, vii.). Nevertheless, I have risked being foolish here because I believe there is one view on free will which stands out as being less unbelievable than the others. The modest incompatibilist position in question has been ably defended by others before me, but without the stress on luck. I have argued that those who defend modest incompatibilism should be honest about the exposure to ultimate luck and that the view is made *more* defensible as a result rather than less.

Bibliography

Adams, R. M. 1977: "Middle Knowledge and the Problem of Evil", *American Philosophical Quarterly,* XIV, 2 pp. 109-117.

Albritton, R. "Freedom of Will and Freedom of Action", in Watson, G. (ed.), *Free Will.* Oxford: Oxford University Press.

Aristotle 1976: *Nicomachean Ethics*, translated by Thomson, J. A. K. Harmondsworth: Penguin.

Arneson, R. 1994: "Autonomy and Preference Formation", in Coleman, J. L. and Buchanan, A., (eds.), *In Harm's Way.* Cambridge: Cambridge University Press.

Austin, J. L. 1961: *Philosophical Papers.* Oxford: Oxford University Press

Ayer, A. J. 1954: *Philosophical Essays.* New York: St Martin's Press.

——— 1973: *The Central Questions of Philosophy.* London: Penguin Books.

Balaguer, M. 2004: "A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will". *Nous,* pp. 379-406.

Bennett, J. 1980: "Accountability", in van Straaten, Z. (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson.* Oxford: Oxford University Press.

Benson, P. 1987: "Freedom and Value", *Journal of Philosophy* 84, pp. 465-487.

Berlin, I. 1969: *Four Essays on Liberty.* Oxford: Oxford University Press.

Bernstein, M. 1989: "Review of R. Kane, Free Will and Values". *Nous* 23, pp. 557-559.

——— 1995: "Kanean Libertarianism", *Southwest Philosophy Review* 11, pp. 151-157.

Berofsky, B. 1987: *Freedom from Necessity.* London: Routledge & Kegan Paul.

——— 2000: "Ultimate Responsibility in a Determined World", *Philosophy and Phenomenological Research* 60. pp. 135-140

——— 2002: "Ifs, Cans, and Free Will: The Issues", in Kane, R. (ed.), The Oxford Handbook of Free Will. Oxford: Oxford University Press.

Blumenfeld, D. 1971: "The Principle of Alternate Possibilities", *Journal of Philosophy* 67, pp. 339-44.

Bohm, D. 1984: *Causality and Chance in Modern Physics.* London: Routledge

Bok, H. 1998: *Freedom and Responsibility.* Princeton: Princeton University Press.

——— 2003: "Freedom and Practical Reason", in Watson, G. (ed.), *Free Will.* Oxford: Oxford University Press.

Borges, J. L. 1974: *Fictions.* London: Calder and Boyars.

Bramhall, J. 1844: *The Works of John Bramhall.* Oxford: John Henry Parker.

Broadie, S. 1991: *Ethics with Aristotle.* Oxford: Oxford University Press.

Carter, W. R. 1979: "On Transworld Event Identity", *Philosophical Review* 88, pp. 443-452.

Chisholm, R. 1976: *Person and Object.* LaSalle, IL: Open Court.

Clarke, R. 1993: "Toward a Credible Agent-Causal Account of Free Will", *Nous* 27, pp. 191-203.

——— 1995: "Indeterminism and Control", *American Philosophical Quarterly* 32, pp. 125-138.

——— 1996: "Contrastive Rational Explanation of Free Choice", *The Philosophical Quarterly* 46, pp. 185-201

——— 2002: "Libertarian Views: Critical Survey of Noncausal and Event-Causal Accounts of Free Agency", in Kane, R. (ed.), The Oxford Handbook of Free Will. Oxford: Oxford University Press.

——— 2003: *Libertarian Accounts of Free Will. Oxford*: Oxford University Press.

Copp, D. 1997. "Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility", Nous 31, pp. 441-56.

Davis, L. 1979: *Theory of Action.* Englewood Cliffs, N.J.: Prentice Hall.

Della Rocca, M. 1998: "Frankfurt, Fischer and Flickers", *Nous* 32, pp. 99-105.

Dennett, D. 1978: *Brainstorms.* Cambridge, Mass: MIT Press.

——— 1984: *Elbow Room.* Cambridge: Cambridge University Press.

Descartes, R. 1976: *Conversations with Burman,* translated by Cottingham, J. Oxford: Clarendon Press.

Double, R. 1989: "Puppeteers, Hypnotists, and Neurosurgeons", *Philosophical Studies* 56, pp. 163-173.

——— 1991: *The Non-Reality of Free Will.* New York: Oxford University Press.

——— 1996: *Metaphilosophy and Free Will.* Oxford: Oxford University Press.

——— 2002: "The Moral Hardness of Libertarians", *Philo,* Vol. 5, No. 2, pp. 226-234.

——— 2005: "Blaming the Victim and Blaming the Culprit", *Think* No. 10, pp. 21-24

Duff, R. A. 1996: *Criminal Attempts.* Oxford: Oxford University Press.

Dworkin, G. 1970: "Acting Freely", *Nous* 4, pp. 367-383.

——— 1988: *The Theory and Practice of Autonomy.* Cambridge: Cambridge University Press.

Ekstrom, L. 1998: "Protecting Incompatibilist Freedom", *American Philosophical Quarterly* 35, pp. 281-291.

——— 2000: *Free Will: A Philosophical Study.* Boulder, CO: Westview Press.

——— 2002: "Libertarianism and Frankfurt-Style Cases", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Feinberg, J. 1986: *Harm to Self.* New York: Oxford University Press.

Finagrette, H. 1972: *The Meaning of Criminal Insanity.* Berkeley: University of California Press.

Finch, A & Warfield, T. 1998: "The Mind Argument and Libertarianism", *Mind* 107, pp. 515-528.

Fine, G. 1981: "Aristotle on Determinism", *The Philosophical Review* 90, pp. 561-579.

Fischer, J. M. & Ravizza, M. 1998: *Responsibility and Control: A Theory of Moral Responsibility.* Cambridge: Cambridge University Press.

Fischer, J. M. 1982: "Responsibility and Control", *The Journal of Philosophy* 79.1, pp. 24-40.

——— 1994: *The Metaphysics of Free Will.* Oxford: Blackwell.

——— 1995: "Libertarianism and Avoidability: A Reply to Widerker", *Faith and Philosophy* 12, pp. 119-125.

——— 1999a: "Recent Work on Moral Responsibility", *Ethics* 110, pp. 93-139.

——— 1999b: "Alfred R. Mele, Autonomous Agents: From Self-Control to Autonomy", *Nous* 33, pp. 133-143.

——— 2002: "Frankfurt-Type Examples and Semi-Compatibilism", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

——— 2003a. "Ought Implies Can', Causal Determinism and Moral Responsibility", *Analysis* 63, pp. 244-50.

——— 2003b. "Frankfurt-style Compatibilism", Watson, G. (ed.), *Free Will.* Oxford: Oxford University Press.

——— (Forthcoming a): "Free Will and Moral Responsibility" in Copp, D. (ed.) *Oxford Handbook of Ethical Theory.* New York: Oxford University Press.

——— (Forthcoming b): "The Cards that Are Dealt You", *Ethics.*

Foot, P. 1966: "Free Will as Involving Determinism", in Berofsky, B. (ed.), *Free Will and Determinism.* New York: Harper and Row.

Frankfurt, H. 1969. "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy* 66, pp. 829-39.

——— 1988: *The Importance of What We Care About.* Cambridge: Cambridge University Press.

——— 2002: "Reply to John Martin Fischer", in Buss, S & Overton, L. (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge, MIT Press.

Gallois, A. 1977: "Van Inwagen on Free Will and Determinism", *Philosophical Studies* 32, pp. 99-105.

Gert, B. & Duggan, T. 1979: "Free Will as the Ability to Will", *Nous* 13, pp. 197-217.

Ginet, C. 1966: "Might We Have No Choice?", in Lehrer, K. (ed.), *Freedom and Determinism.* New York: Random House.

——— 1983: "In Defence of Incompatibilism", *Philosophical Studies* 44, pp. 391-400.

——— 1990: *On Action*. Cambridge: Cambridge University Press.

——— 1996: "In Defence of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing", *Philosophical Perspectives* 10, pp. 403-17.

Goetz, S. 1997: "Libertarian Choice", *Faith and Philosophy* 14, pp. 195-211.

——— 2005 (forthcoming): "Frankfurt-Style Counterexamples and Begging the Question", *Midwest Studies in Philosophy.*

Haji, I. 1998: *Moral Appraisability*. New York: Oxford University Press.

——— 1999a: "Moral Anchors and Control", *Canadian Journal of Philosophy* 29, pp. 175-203.

——— 1999b: "Indeterminism and Frankfurt-style Examples", *Philosophical Explorations* 2, pp. 42-58.

——— 2002: *Deontic Morality and Control*. New York: Cambridge University Press.

Haji, I. and McKenna, M. 2004: "Dialectical Delicacies in the Debate About Freedom and Alternative Possibilities", *Journal of Philosophy,* Vol. 4, pp. 299-314.

Harrison, G. K. 2005: "Moral Responsibility and the Principle of Avoidable Blame", *Ethic@.*

——— (forthcoming): "Frankfurt-style cases and Improbable Alternative Possibilities", *Philosophical Studies*.

——— (forthcoming): "Frankfurt-Style Cases and the Question Begging Accusation", *Facta Philosophica.*

——— (forthcoming): "Moral Luck and Hyper Libertarianism", *Sorites.*

Hobart, R. E. 1934: "Free Will as Involving Determinism and Inconceivable Without It", *Mind* 43, pp. 1-27.

Hobbes, T. 1651. 1996: *Leviathan* in Morgan, M. (ed.). Cambridge: Hackett Publishing Company.

——— 1840: *The English Works of Thomas Hobbes.* Molesworth, W. (ed.). London: John Bohn.

Hodgson, D. 1999: "Hume's Mistake", in Libet, B., & Freeman, A., & Sutherland, K. (eds.), *The Volitional Brain: Towards a Neuroscience of Free Will.* Thoverten: Imprint Academic.

——— 2002: "Quantum Physics, Consciousness, and Free Will", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Hoffman, J. and Rosenkrantz, G. 2002: *Exploring the Divine Attributes.* Malden, MA: Blackwell.

Honderich, T. 1988: *A Theory of Determinism.* 2 Vols. Oxford: Clarendon Press

——— 2002: "Determinism and the Real Problem", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Horgan, T. 1985: "Compatibilism and the Consequence Argument", *Philosophical Studies* 47, pp. 339-356.

Hume, D. 1955: *An Enquiry Concerning Human Understanding* Indianapolis: Bobbs-Merrill.

Hunt, D. 2000: "Moral Responsibility and Unavoidable Action", *Philosophical Studies* 97.2, pp. 195-227.

Irwin, T. 1980: "Reason and Responsibility in Aristotle", in Rorty, A. (ed.), *Essays on Aristotle's Ethics.* Berkeley: University of California Press.

Kane, R. 1985: *Free Will and Values* Albany: State University of New York Press: New York.

——— 1988: "Libertarianism and Rationality Revisited", *The Southern Journal of Philosophy* 26, pp. 441-460.

——— 1996: *The Significance of Free Will.* New York: Oxford University Press.

——— 1999: "On Free Will, Responsibility, and Indeterminism: Responses to Clarke, Haji, and Mele", *Philosophical Explorations* 2, pp. 105-121.

——— 2002: "Some Neglected Pathways in the Free Will Labyrinth", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

——— 2003: "Responsibility, Luck and Chance", in Watson, G. (ed.), *Free Will*, Oxford: Oxford University Press.

——— 2005: "Commentary", presented at the *31st Annual Meeting of The Society of Philosophy and Psychology, June 2005,* and published on the Garden of Forking Paths website.

Kant, I. 1960: *Religion within the Bounds of Reason Alone*, translated by Greene, T. and Hudson, H. New York: Harper and Row.

Kapitan, T. 1989: "Doxastic Freedom: A Compatibilist Alternative", *American Philosophical Quarterly* 26, pp. 31-41.

——— 2002: "A Master Argument for Incompatibilism?", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Klein, M. 1990: *Determinism, Blameworthiness, and Deprivation.* Oxford: Oxford University Press.

Latus, A. 2001: "Moral and Epistemic Luck", *Journal of Philosophical Research* 25, pp. 149-172.

Lehrer, K. 1966: "An Empirical Disproof of Determinism", in (ed.), Lehrer, K. *Freedom and Determinism.* New York: Random House.

Lewis, D. 2003: "Are We Free to Break the Laws?", in Watson, G. (ed.), *Free Will.* Oxford: Oxford University Press.

Locke, J. 1965: *An Essay Concerning Human Understanding.* Cranston, M. (ed.). New York: Collier.

McCann, H. 1998: *The Works of Agency: On Human Action, Will and Freedom.* Ithaca, NY: Cornell University Press.

McCloskey, H. J. 1965: "A Non-utilitarian Approach to Punishment", *Inquiry* 8, pp. 249-263.

——— 1967: "Utilitarianism and Retributive Punishment", *Journal of Philosophy,* 64, pp. 91-110.

——— 1973: "Utilitarianism: Two Difficulties", *Philosophical Studies* 24, pp. 62-63.

McKenna, M. 1997: "Alternative Possibilities and the Failure of the Counterexample Strategy", *Journal of Social Philosophy* 28, pp. 71-85.

Mele, A. 1995: *Autonomous Agents: From Self-Control to Autonomy.* Oxford: Oxford University Press.

——— 1998: "Review of Robert Kane, 'The Significance of Free Will'", *Journal of Philosophy* 95, pp. 581-584.

——— 1999a: "Kane, Luck and the Significance of Free Will", *Philosophical Explorations* 2, pp. 96-104.

——— 1999b: "Ultimate Responsibility and Dumb Luck", *Social Philosophy and Policy* 16, pp. 274-293.

——— 2002: "Autonomy, Self-Control, and Weakness Of Will", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Mele, A. and Robb 1998: "Rescuing Frankfurt-Style Cases", *The Philosophical Review* 107.1, pp. 97-112.

——— 2003: "BBs, Magnets, and Seesaws: The Metaphysics of Frankfurt-Style Cases", in Widerker, D. & McKenna, M. (eds.), *Moral Responsibility and Alternative Possibilities.* Aldershot: Ashgate.

Mill, J. S. 1958: *Autobiography.* London: Oxford University Press.

Moore, G. E. 1962: *Philosophical Papers.* New York.

——— 1993: *Principia Ethica.* Cambridge: Cambridge University Press.

Nagel, T. 1979: *Mortal Questions.* Cambridge: Cambridge University Press

Narveson, J. 1977: "Compatibilism Defended", *Philosophical Studies* 32, pp. 83-88.

Neely, W. 1974: "Freedom and Desire", *Philosophical Review* 83, pp. 32-54.

Nietzsche, F. 1996: *Beyond Good and Evil*, translated by Kaufmann, W. reproduced in Morgan, M. (ed.), *Classics of Moral and Political Theory.* Cambridge: Hackett Publishing Company.

Nozick, R. 1981: *Philosophical Explanations.* Cambridge, Mass: Harvard University Press.

O'Connor, T. 2000: *Persons and Causes.* New York: Oxford University Press.

——— 2002: "Libertarian Views: Dualist and Agent-Causal Theories", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

——— 2003: "Agent Causation", in Watson, G. (ed.), *Free Will*, Oxford: Oxford University Press.

Otsuka, M. 1998: "Incompatibilism and the Avoidability of Blame", *Ethics* 108, pp. 685-701.

Peacocke, C. 1999: *Being Known.* Oxford: Clarendon Press.

Pereboom, D. 1995: "Determinism al Dente", *Nous* 29, pp. 21-45.

——— 2000: "Alternate Possibilities and Causal Histories", *Philosophical Perspectives* 14, pp. 119-138.

——— 2001: *Living Without Free Will,* Cambridge: Cambridge University Press.

——— 2002: "Living Without Free Will: The Case for Hard Incompatibilism", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Perry, D. 1965: "Prediction, Explanation, and Freedom", *Monist* 49/2.

Pike, N. 1965: "Divine Omniscience and Voluntary Action", *Philosophical Review* 74, pp. 27-46.

Plantinga, A. 1974: *The Nature of Necessity,* New York: Oxford.

Powys, J. 1974: *The Art of Happiness.* London: Village Press.

Ravizza, M. 1994: "Semi-Compatibilism and the Transfer of Non-Responsibility", *Philosophical Studies* 75, pp. 61-93.

Rescher, N. 1993: "Moral Luck", in Statman, D. (ed.), *Moral Luck.* Albany: State University of New York Press.

Richards, N. 1993: "Luck and Desert", in Statman, D. (ed.), *Moral Luck.* Albany: State University of New York Press.

Rosebury, B. 1995: "Moral Responsibility and Moral Luck", *Philosophical Review* 104, pp. 499-524

Rowe, William L. 2000: "The Metaphysics of Freedom: Reid's Theory of Agent Causation", *American Catholic Philosophical Quarterly* 74, pp. 425-46.

Russell, P. 2002a: "Critical Notice", *Canadian Journal of Philosophy* 32.4, pp. 587-606.

——— 2002b: "Pessimists, Pollyannas, and the New Compatibilism", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Saunders, J. 1968: "The Temptations of Powerlessness", *American Philosophical Quarterly* 5, pp. 100-108.

Schlick, M. 1939: *Problems of Ethics.* New York: Dover.

Schlossberger, E. 1986: "Why We Are Responsible for Our Emotions", *Mind,* Vol. 95, No. 377, pp. 37-56.

——— 1992: *Moral Responsibility and Persons.* Temple: Philadelphia.

Schnall, I. 2001: "The Principle of Alternate Possibilities and 'ought' implies 'can'", Analysis 61, pp. 335-40.

Schoeman, F. 1978: "Responsibility and the Problem of Induced Desires", *Philosophical Studies* 34, pp. 293-301.

——— 1979: "On Incapacitating the Dangerous", *American Philosophical Quarterly* 16, pp. 56-67.

Sennett, J. 1999: "Is There Freedom in Heaven", *Faith and Philosophy,* Vol. 16, No. 1, pp. 69-82.

Sidgwick, H. 1963: *Methods of Ethics.* London: Macmillan.

Sinnott-Armstrong, W. 1988: *Moral Dilemmas.* Oxford: Blackwell Publishers.

Smart, J. J. C. 1973: "An Outline of a System of Utilitarian Ethics", in *Utilitarianism For and Against.* Cambridge: Cambridge University Press.

Smilansky, S. 1994: "The Ethical Advantages of Hard Determinism", *Philosophy and Phenomenological Research* 54, pp. 355-363.

——— 2000: *Free Will and Illusion.* Oxford: Oxford University Press.

——— 2002: "Free Will, Fundamental Dualism, Centrality of Illusion", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

——— 2004: "Fischer on Free Will and Luck", presented at the American Philosophical Association Central Division Meetings, April 2004.

Sorabji, R. 1980: *Necessity, Cause, and Blame: Perspectives on Aristotle's Philosophy* Ithaca, N.Y.: Cornell University Press.

Sprigge, T. L. S. 1965: "A Utilitarian Reply to Dr. McCloskey", *Inquiry* 8, pp. 264-291.

Squires, J. 1969: "Blame", in Acton, H. B. (ed.), *The Philosophy of Punishment.* London: Macmillan.

Statman, D. 1993: "Introduction", in *Moral Luck.* Albany: State University of New York Press.

Stevenson, C. L.: 1944: *Ethics and Language.* New Haven and London: Yale University Press.

Strawson, G. 1986: *Freedom and Belief.* Oxford: Clarendon Press.

——— 2000: "The Unhelpfulness of Indeterminism", *Philosophy and Phenomenological Research* 60, pp. 149-155.

——— 2003: "The Impossibility of Moral Responsibility", in Watson, G. (ed.), *Free Will.* Oxford: Oxford University Press.

Strawson, P. 1980: "Reply to Ayer and Bennett", in van Straaten, Z. (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson.* Oxford: Oxford University Press.

——— 2003: "Freedom and Resentment", in Watson, G. (ed.), *Free Will*, Oxford: Oxford University Press.

Stump, E. 1990: "Intellect, Will, and the Principle of Alternative Possibilities", in Beaty, M. (ed.), *Christian Theism and the Problems of Philosophy.* Notre Dame: University of Notre Dame Press.

——— 1995: "Cartesian Dualism and Materialism Without Reductionism", *Faith and Philosophy* 12, pp. 505-531.

——— 1996: "Libertarian Freedom and the Principle of Alternative Possibilities", in Howard-Snyder, D. and Jordan, J. (eds.) *Faith, Freedom and Rationality.* Lanham, MD: Rowman and Littlefield.

——— 1999: "Dust, Determinism, and Frankfurt: A Reply to Goetz", *Faith and Philosophy,* Vol. 16, No. 3, pp. 413-422.

Sverdlick, S. 1993: "Crime and Moral Luck", in Statman, D. (ed.), *Moral Luck.* Albany: State University of New York Press.

Taylor, C. & Dennett, D. 2002: "Who's Afraid of Determinism?" in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Taylor, R. 1966: *Action and Purpose.* Englewood Cliffs, NJ: Prentice-Hall.

——— 1976: "Responsibility for Self", in Rorty, A. (ed.), *Identities of Persons.* Berkeley: University of California Press.

——— 1992: *Metaphysics*, 4th edition. Englewood Cliffs: Prentice-Hall.

Thomson, J. J. 1993: "Morality and Bad Luck", in Statman, D. (ed.), *Moral Luck.* Albany: State University of New York Press.

Unger, P. 1984: *Philosophical Relativity.* Minneapolis: University of Minnesota Press

——— 1986: "The Cone Model of Knowledge", *Philosophical Topics* vol. 14, pp. 125-178

Van Inwagen, P. 1983: *An Essay on Free Will.* Oxford: Clarendon Press.

——— 1989: "When Is The Will Free?" in Tomberlin, J. (ed.), *Philosophical Perspectives*, Vol. 3. Atascadero: Ridgeview.

——— 1997: "Against Middle Knowledge", *Midwest Studies in Philosophy* 21, pp. 50-69

——— 2002: "Free Will Remains a Mystery" in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Velleman, J. D 1989: *Practical Reflection.* Princeton, N. J.: Princeton University Press.

Vihvelin, K. 1991: "Freedom, Causation, and Counterfactuals", *Philosophical Studies* 64, pp. 161-184.

Waller, B. 1988: "Free Will Gone Out of Control: A Review of R. Kane's *Free Will and Values", Behaviourism* 16, pp. 149-162.

——— 1990: *Freedom Without Responsibility.* Philadelphia: Temple University Press.

Walter, E. 1978: "Is Libertarianism Logically Coherent?", *Philosophy and Phenomenological Research* 38, pp. 505-13.

Warfield, T. 1996: "Determinism and Moral Responsibility Are Incompatible", *Philosophical Topics* 24, pp. 215-226

Watson, G. 1975: "Free Agency", *Journal of Philosophy* 72, pp. 205-220.

——— 1987: "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme", in Shoeman, F. (ed.), *Responsibility Character, and Emotions.* Cambridge: Cambridge University Press.

——— 1996: "Two Faces of Responsibility", *Philosophical Topics* 24, pp. 227-248.

——— 2003: "Introduction", in Watson, G. (ed.), *Free Will.* Oxford: Oxford University Press.

Widerker, D. 1987: "On an Argument for Incompatibilism", *Analysis* 47, pp. 37-41.

——— 1991: "Frankfurt on 'Ought Implies Can' and Alternative Possibilities", Analysis 51, pp. 222-24.

——— 1995: "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities", *Philosophical Review* 104, pp. 247-61.

——— 2000: "Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look", *Philosophical Perspectives* 14, pp. 181-201.

——— 2002: "Responsibility and Frankfurt-Type Examples", in Kane, R. (ed.), *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Wiggins, D. 1973: "Towards a Reasonable Libertarianism", in Honderich, T. (ed.), *Essays on Freedom of Action.* London: Routledge & Kegan Paul.

Williams, B. 1981: *Moral Luck.* Cambridge: Cambridge University Press.

Wolf, S. 1990: *Freedom Within Reason.* New York: Oxford University Press.

——— 2003: "Sanity and the Metaphysics of Responsibility", in Watson, G. (ed.), *Free Will.* Oxford: Oxford University Press.

Wyma, K. 1997: "Moral Responsibility and Leeway for Action", *American Philosophical Quarterly* 34, pp. 57-70.

Yaffe, G. 1999: "Ought Implies Can' and the Principle of Alternate Possibilities", *Analysis* 59, pp. 218-22.

Young, R. 1979: "Compatibilism and Conditioning", *Nous* 13, pp. 361-378.

Zagzebski, L. 2000: "Does Libertarian Freedom Require Alternative Possibilities?", *Philosophical Perspectives* XIV, pp. 231-248.

Zimmerman, D. 1981: "Hierarchical Motivation and the freedom of the Will", *Pacific Philosophical Quarterly* 62, pp. 354-368.

——— 2000: "Making Do: Troubling Stoic Tendencies in an Otherwise Compelling Theory of Autonomy", *Canadian Journal of Philosophy* 30, pp. 25-53.

Zimmerman, M. 1989: *An Essay on Moral Responsibility.* Totowa, N.J.: Rowman and Littlefield.

——— 1993: "Luck and Moral Responsibility", in Statman, D. (ed.), *Moral Luck.* Albany: State University of New York Press.

——— 2002: "Taking Luck Seriously", *The Journal of Philosophy,* Vol. XCIX, No. 11, pp. 553- 576.