FREE WILL AND RATIONALITY

A DEFENSE OF THE VIEW THAT FREE AND RESPONSIBLE AGENTS CAN PERFORM ONLY THE RIGHT ACTIONS FOR THE RIGHT REASONS

by Damir Cicic

Submitted to

Central European University
Department of Philosophy

In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Philosophy

Supervisor: Professor Ferenc Huoranszki

Budapest, Hungary

2015

Abstract

This dissertation offers an unorthodox answer to the two main questions in the free will debate – the question how is free will as a condition of moral responsibility possible, and the question whether we actually have it. It suggests that free will is possible and that we have it only if it consists in the ability to do right things for the right reasons and if that ability cannot be unexercised. In other words, this dissertation suggests that the only free actions are the right actions performed for the right reasons. This suggestion is based on considerations of the main the main skeptical challenges to free will and on Susan Wolf's account of free will. The first chapter, deals with the main challenge to the claim that ability to do otherwise exist if determinism is true - the so called Consequence Argument - and concludes that the argument is very plausible. In the second chapter, an argument suggested by Harry Frankfurt to the effect that the Consequence Argument is irrelevant because free will does not involve ability to do otherwise is considered and rejected. The third chapter focuses on two objections to libertarian theories of free will - the objection that indeterminism undermines free will by undermining control, and objection that indeterminism is irrelevant because it does not provide more space for control than determinism. These objections are rejected but it is shown that the only version of libertarianism which avoids them is not very attractive. The fourth chapter defends Susan Wolf's view and the thesis that free will is asymmetric which her view entails. In addition, it suggests that her view can be defended more easily if the possibility of misuse of free will is excluded. The final chapter shows that the proponent of Wolf's view must exclude this possibility in order to defend compatibilism about free will and determinism from the 'manipulation arguments.' It also shows that impossibility of free wrongdoing follows from the acceptance of asymmetry of Wolf's view and incompatibilism about the ability to do otherwise and determinism.

Acknowledgements

I would like to express my gratitude to my advisor Professor Ferenc Huoranszki for many conversations which deepened my interest in free will and comments which helped me to develop my own view. I would also like to thank Professor Gary Watson for encouraging me to pursue my own interests and for being very gracious with his time in discussing the problem of free will with me. I want to thank also Professor Hanoch Ben-Yami and Professor Michael Griffin for their continuous support during my studies, as well as my friends from PhD program Hywel Griffiths, Anton Markoc, Mojca Kuplen, Isik Sarihan, and my friends from undergraduate studies in Belgrade Arandjel Bojanovic, Goran Boroja, Bozidar Maslac and Dusko Majkic for spending a lot of time discussing free will with me and for showing interest in my ideas. Finally, I would like to thank my mother Ljiljana Rebronja and my father Behudin Cicic for their loving support and my grandparents Budimir and Franciska Djordjevic and for giving me the opportunity to focus on my dissertation.

Contents

INTRODUCTION

Free will is no doubt one of the greatest mysteries of human nature. On one hand, it is one of our most valuable assets. Without it many things that make our lives worth living would not be available to us. It is also the source of dignity that other creatures in nature lack. Without free will our actions would not be truly attributable to us; we would not be morally responsible for what we do; it would perhaps not make sense to talk about love, compassion, friendship, or even about morality; there would be no real creativity in the world; none of our actions would be significantly different from actions resulting from various pathological states such as addictions, phobias or manias. In other words, without free will our lives would not make sense, and we would not even be humans. Yet, it is very difficult to prove that we have free will or even that free will could exist. But what is most puzzling, we don't really know what it is.

Of course, we have some general ideas about what it should be. Thus, we know that it has something to do with control over our own behavior, especially our own decisions and actions or omissions resulting from our decisions. Or more precisely, it is a sort of power that underlies control involved in these forms of behavior. This is clear because people who lack free will, like people in the above mentioned pathological states, seem to lack that sort of control. In addition, it is clear that it is a power that can be exercised consciously in the light of considerations that speak for or against certain courses of action. For whatever free will is, it would not be of much value if it were impossible to exercise it in such a way; it would not be a power that we ascribe to adult human beings, which distinguishes them from other beings and which makes living meaningful.

But, what kind of power is free will? No doubt the most natural answer to this question is that it is a power to perform or not to perform certain actions and to will or not

will to perform those actions. In other words, the most natural answer is that it is a power that involves alternative possibilities or ability to do otherwise. It is often assumed that when one has this power one also has the power to be the author or an appropriate source of one's actions and vice versa. However, some philosophers think that these two 'powers' are not just different aspects of one power, but really different powers that do not necessarily occur together. In particular, they believe that one may be an appropriate source of action without having the ability to do otherwise. It is common now to refer to the two forms of control grounded in these two powers (using terminology established by John Martin Fischer) as regulative control – control which involves alternative possibilities- and guidance for control which does not involve alternative possibilities.

But there is also a third answer which identifies free will neither with a power to 'regulate' nor to 'guide' behavior but in the first place power to do certain concrete things. Thus, some philosophers believe that free will can be understood in terms of the power to critically examine one's reasons and act in accordance and on the basis of those reasons, or to act in accordance with one's values. These powers may or may not involve ability to do otherwise depending on how they are understood.

There are significant differences between the particular theories of free will which fall into these categories as well as significant similarities between the theories in different categories. For instance, among those who think that free will essentially involves ability to do otherwise some believe that free will requires the falsity of determinism – the thesis that everything that happens is a necessary consequence of what happened in the past and the laws of nature. These philosophers are called incompatibilists; those who disagree with them are compatibilists. Among the former some – the so called libertarians - believe that free will is possible if indeterminism is true. Others - the so called impossibilists or hard incompatibilists – think that free will impossible. The parallel divisions exist among those

who think that free will essentially grounds guidance control over one's own behavior. In addition, philosophers in this group are divided on those who think that free will is a historical phenomenon and those who think that it depends only on the agent's properties at the time of action. On the other hand, the similarity between the theories of those who think of free will in terms of guidance control and those who think in terms of some concrete powers is that they usually put emphasis on the importance of agent's reasons or rationality for free will. Finally, conceptions of free will of particular philosophers also vary relative to the value that free will is supposed to secure. Thus, some think that freedom involving ability to do otherwise might be relevant for some purposes, but not when it comes to the question of moral responsibility.

However, there is an assumption that all contemporary philosophers accept, regardless of their favored conception of free will. It is the assumption that free will (if it can exist) can be exercised for virtually any purpose. In other words, they accept, or rather just presuppose that free will can be exercised for a good or for a bad purpose, i.e. that it can be a property of both heroes and the villains.

At first sight, this is a totally innocent assumption. However, the fact that the possibility of akrasia or the free action against one's better judgment represents a philosophical problem and that some great philosophers such as Socrates, R.M. Hare and Gary Watson endorsed skepticism about this phenomenon shows that it is not meaningless to ask whether this assumption really is true. More precisely, this is so if we assume, as I believe we should, that a directly free bad action can only be an akratic action. I will give later (in chapter 5) a more detail explanation of why I think that this is so. Here it will suffice, I think, to say that it is intuitively true that there is no real viciousness without awareness that one is acting viciously. Also, it will suffice to say that direct freedom is freedom which is not derived from freedom at some earlier time.

As the title of my dissertation suggests, I also believe that we have good reasons to think that akratic action is impossible. However, I have not come to this conclusion by considering the nature of akratic actions. I have come to this conclusion by trying to answer the question: is free will possible? The answer to which my research has lead me is: yes, but only if we reject the possibility of free actions performed against one's conception of what is the right thing to do. In other words, my answer to this question is that free will is possible, but only when we do the right things for the right reasons.

My interest in this question has source in my puzzlement over the skeptical challenge, presented first by William James, known as the 'dilemma argument.' This challenge basically goes like this:

1) If determinism is true, no one can have free will.
2) If indeterminism is true no one can have free will.
3) Either determinism or indeterminism must be true.

Therefore,

4) Free will is impossible.

In my view, this argument is a serious threat to the possibility of free will. For, it is obviously valid and we have very good reasons to think that its premises are true. There are two reasons for thinking that the first premise of this argument is true. The first reason concerns those who think that free will involves ability to do otherwise. It says that if determinism is true no one can have free will because in that case no one can have ability to do otherwise. The main argument for the claim that if determinism is true no one has the

ability to do otherwise is called the Consequence Argument. I discuss this argument in the first chapter and conclude that it is sound.

The other reason for thinking that free will is incompatible with determinism concerns mainly those who do not think that free will requires ability to do otherwise, but think that one has free will only if one can be the (appropriate) source of one's actions. For, some philosophers think that if determinism is true no one can be the appropriate source of one's actions. The main argument for this claim is the so called Manipulation Argument. I consider this argument in the fifth chapter and reject it.

The main argument for the second premise is that indeterminism entails chance which in turn entails that all actions which are undetermined cannot be free because what is a matter of chance cannot be under anyone's control. In the third chapter I consider and reject this claim. However, in the same chapter I consider another claim about indeterminism that can replace this premise in the argument. It is the claim that indeterminism does not provide more space for free will than determinism. I conclude that it is very plausible to think that this claim is true, although to think otherwise does not seem incoherent.

My answers to the premises of this argument leave open two ordinary ways of defending the possibility of free will that are open to me which don't require the above mentioned claim that we cannot act freely when we do bad things. They leave open the possibility to argue that free will is possible because it does not require ability to do otherwise i.e. that it provides guidance control, or to argue that free will is possible only if indeterminism is true. However, in addition to the obvious problem that these two answers seem to clash with each other, there is also a problem that I don't find either of these conceptions of free will promising for independent reasons.

There are two reasons why I don't find the former answer promising. First, I don't see a reason to claim that ability to do otherwise is never required for free will. Many

philosophers accept this claim because of an argument based on the so called Frankfurt-style examples which I consider and reject in chapter 2. Second, I think that we have very good reasons to think that for free will be possible when we do bad things intentionally, it must involve ability to do otherwise. I explain why I think so in chapter 4.

On the other hand, the idea that free will is possible only if indeterminism is true does not seem to me as a solution for three reasons. First, as I mentioned above, I don't see determinism as a problem for origination. That is, I don't think that we cannot be appropriate sources of our actions if determinism is true. Second, I think that there are good reasons for thinking that ability to do otherwise is not necessary for acting with free will in some cases. I explain this in chapter 4. Finally, as I said above, although I think the idea that indeterminism provides space for more freedom is not incoherent, I think that it is very plausible to think that it is not so.

Therefore, none of the existing defenses of the possibility free will against the dilemma argument seems successful to me. But, how can the rejection of the possibility of exercise of free will for bad purposes help? To see that we must focus on the notion of free will as a specific ability to do good things for good reasons suggested by Susan Wolf. For, in my view, this notion of free will is immune to all of the problems that we encounter in arguing for other views, except of some that I will discuss at the end of chapter 4 and in chapter 5 which can be eliminated only by assuming that the ability which is essential for this motion of free will cannot be unexercised.

Of course, most readers would not agree with me because the impossibility of explaining the possibility of ability to freely do bad things may seem to be a great problem, bigger perhaps than other problems. One might say that this solution amounts to throwing a baby out together with bathing water. One might say that such a mutilated free will cannot give us things that for which we value free will. In particular, one might say that this notion

of free will is not satisfactory when it comes to explaining moral responsibility. This would be a problem for my argument because like most philosophers, I am interested in free will mainly because of moral responsibility. Moreover, I define free will as a sort of control over one's own behavior which is necessary for moral responsibility.

However, I don't see any reason to think that the notion of free will that I suggest should not be relevant for moral responsibility. On the standard view, moral responsibility is a property in virtue of which a person deserves blame for wrongdoings or praise for behaving in the right way or doing something good. If it is impossible to do bad things freely, then no one ever deserves blame for anything. However, I don't see why should that entail that no one ever deserves praise for anything if doing good things freely is possible and other requirements for moral responsibility are satisfied.

I admit, however, that my thesis is revisionary and at first sight counterintuitive. I am aware that the belief that we are sometimes blameworthy and even deserve punishment for what we do is natural and that it might be impossible to reject it completely. However, even if this is true, I think that we should not reject the possibility that the belief in question is false. For, our feelings are not the best tool for bringing us closer to the truth. They often obscure our vision rather than sharpening it. In what follows I will argue that we have very good reasons to think that they do this to our understanding of free will and moral responsibility, at least when responsibility for wrong actions is concerned.

CHAPTER 1: DETERMINISM AND ABILITY TO DO OTHERWISE

The question about the compatibility of free will and determinism is no doubt the most popular question concerning free will. Arguably, the main reason for this is that it is very easy to see why determinism constitutes a threat to free will. Determinism is the thesis that every event is a necessary consequence of antecedent events and the laws of nature. In other words, determinism implies that only what actually happens is possible to happen. But if free will involves ability to do otherwise, determinism and free will seem to be incompatible. For, it is difficult to see how one can be able to do otherwise if doing otherwise is impossible.

It is not surprising for that reason that upon the first encounter with this topic most people find compatibilism very puzzling. The claim that free will and determinism are incompatible seems so obvious that it is easier to believe that many great minds who have thought otherwise have deceived themselves in order to save those things which seem to depend on the existence of free will. However, on closer examination, it becomes clear that there are very good reasons to doubt that the incompatibility thesis is true. Most importantly, closer examination shows that it is very difficult to establish the connection between the sense in which our actions are necessary if determinism is true and the sense in which they are necessary in cases in which we clearly lack free will.

On the basis of this insight many philosophers have concluded that the incompatibility thesis or incompatibilism must rest on some sort of confusion. In particular, they have argued that those who accept this thesis - incompatibilists - have a mistaken conception of determinism, either because of the conflation causal determination and compulsion,

descriptive and prescriptive laws or for some other reason. In other words, they have concluded that determinism initially seems to be a threat to free will only because of our failure to make a difference between different sorts of necessitation which undermine free will necessitation connected to determinism.

However, there seems to be a way of showing that determinism and ability to do otherwise are incompatible that does not rest on any confusion about causation or the laws of nature. It is called the Consequence Argument. Informally, the argument says that since we don't have control over the states of the world before we were born and the laws of nature, and we don't have control over the fact that in deterministic worlds everything that happens, including everything we do, is a consequence of the states of the world before we were born and the laws of nature, in deterministic worlds we do not have control over anything we do.

Several prominent philosophers have offered formal version of this argument and tried to show that it is sound, i.e. that it rests on true premises and uses valid rules of inference. However, their attempts have encountered strong criticisms. According to critics, all versions of the argument fail because there is no unique interpretation of the term in the argument which refers to free will on which both its premises and the rules of inference are true. In other words, the critics of the argument accuse its proponents of committing the fallacy of equivocation.

In this chapter I defend the Consequence Argument from this objection. I focus on two formal versions of the argument that Peter van Inwagen presented in his book *An Essay on Free Will* which are generally regarded as the strongest versions of this argument. The main virtues of his two versions of the argument over other versions are that they use terminology familiar in the free will debate (unlike another argument that van Inwagen presents in the same book) and make the rules of inference used in them very explicit. The main difference between his two versions of the Consequence Argument is that one of them

uses *modal* rules of inference, while the other uses only the rules of ordinary logic. For that reason, I will call the former version 'the modal argument' and the latter 'the non-modal argument' although there is a sense in which both versions are modal in so far as they are about the *ability to do otherwise*, which is a modal concept. I begin by discussing the modal version of the argument because I believe that some insights from that discussion will be useful in the discussion of the non-modal version of the argument.

1.1 The Modal Argument

Van Inwagen's modal argument connects determinism with the capacity to make choices and thereby control things in one's environment. It purports to show that if determinism is true no one has, or ever had, a choice about anything. In his book, van Inwagen does not say what exactly 'not having a choice about something' means, but he says that the expression plays the role of a modal operator in the argument. This means that special modal rules of inference apply to sentences which containing this expression. Using N as a symbol for this phrase, Van Inwagen presents the following inference rules:

Alpha: $\Box P \models NP$ and,
Beta: $Np, N(p \rightarrow q) \models Nq$

Less formally, Alpha says that if something is necessarily true (in the broadly logical sense), then no one has or ever had a choice about that. Beta says that if no one has or ever had a choice about the truth of one proposition and that proposition *materially implies* some other propositions, then no one has or ever had a choice about the truth of that other proposition.

The argument has three premises. One of the premises states that no one has, or ever had a choice about whether the laws of nature (L) are as they are (NL). The other premise

says the same about the relation of any human being to some state of the world (Po) before any human being existed (NPo). Finally, there is a premise which says that if determinism is true, we can deduce any sentence about the present state of the world from the conjunction of the sentence expressing the complete state of the universe at some time in the past when no human being existed (Po) and the sentence expressing the conjunction of all the laws of nature (L).

Van Inwagen presents the argument in the following way:

(1) □(Po & L→ P) premise 1
(2) □(Po→(L→P) by logic from (1)
(3) N(Po→(L→P) by Alpha from (2)
(4) NPo. Premise 2
(5) N(L→P) by Beta from (3) and (4)
(6) NL Premise 3
(7) **N**pby Betafrom (5) and (6)

This argument seems very plausible at first sight. But, as I mentioned above, the main objection to all versions of the Consequence Argument is that they rest on equivocation related to the term capturing the meaning of free will. So, let us check if this objection applies to this version of the argument.

Since the expression 'has a choice about' captures the meaning of the term 'free will' in this argument, we should check if there is a unique interpretation of this expression on which premises of the argument are true and the inference rules valid. More precisely, we must see whether there is such an interpretation of this expression that is relevant to the discussion about free will and determinism.

Consider first the premises of the argument. Premise 1 is clearly irrelevant in the context of this inquiry because it does not contain the expression in question. So, we should focus on premises 2 and 3. These premises are obviously true under any ordinary

interpretation of 'having a choice about.' For, it is obvious that no one has or ever had a power to influence directly or indirectly things that happened before one existed or the laws of nature. In other words, the past before any human being was born and the laws of nature are totally independent of our present powers.

What about the inference rules? The rule Alpha is also obviously true under any ordinary interpretation of 'having a choice about.' For, it is clearly not up to anyone (except perhaps God) what necessary truths there are. However, the rule Beta is a little bit tricky. On one hand, Beta sounds very plausible and it is easy to find arguments in which it leads from true premises to true conclusions. Consider, for example, the following argument presented by van Inwagen:

> The sun will explode in 2000 **AD,** and no one has, or ever had, any choice about whether the sun will explode in 2000 **AD;**
> If the sun explodes in 2000 **AD,** all life on earth will end in 2000 **AD,** and no one has, or ever had, any choice about whether, if the sun explodes in 2000 **AD,** all life on earth will end in 2000 **AD;**
> *hence,* All life on earth will end in 2000 **AD,** and no one has, or ever had, any choice about whether all life on earth will end in 2000 **AD.**[1]

This argument is clearly valid under any interpretation of the expression 'no one has or ever had a choice about.' However, little reflection shows that Beta is not valid under any interpretation of the expression in question. For example, if we take 'having a choice about something' to mean 'being able to causally influence something by choosing,' we encounter invalid arguments resting on Beta. For, it is true that no one has or ever had a choice about the past before anyone was born or about the laws of nature. And it is true that under determinism (given Alpha) no one has or ever had a choice about the fact that the conjunction of propositions about the past and the laws of nature implies what we actually do. But that

[1] Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), 98.

does not entail, as Beta recommends, that no one has or ever had the power to influence anything by one's choices or even to make choices.

So, in order to see whether this version of the Consequence Argument rests on equivocation, we must examine if there is some ordinary interpretation of the operator N, (i.e. expression 'no one has or ever had a choice about') on which the rule Beta is valid. Alternatively, if there is no such interpretation, we must see if there is some non-standard interpretation of this expression on which this rule is valid and the premises of the argument true.

We must keep in mind, however, that the candidate interpretations of the operator N must be relevant for the discussion about free will, and in particular for free will as involving ability to do otherwise. For, the Consequence Argument is not designed to show that if determinism is true no one makes choices (in the psychological sense) or that no one can influence things by their choices, but that *no one can do otherwise* if determinism is true because under determinism no one can make alternative choices. Therefore, we must first check if Beta is valid under some interpretation of N which implies the absence of ability to do otherwise or the idea that what the agent does is unavoidable.

1.1.1 Validity of Beta and Different Interpretations of Operator N

One way to capture the relevant sense of having a choice about the truth of some proposition p is to say that p is true and the person could *ensure* that p is false. (It is plausible to assume that van Inwagen understood in fact the expression in this way). If having a choice is interpreted in this way the plausibility of Beta becomes much easier to appreciate. For, it is difficult to see how one could ensure the falsity of a consequent without the power to ensure either the falsity of the antecedent or the falsity of the implication.

Surprisingly, however, there are clear counterexamples to Beta if (not) having a choice is interpreted in this way. Consider the following case presented by Timothy O'Connor:

> Suppose that Helen is deliberating about whether or not to insult Stewart. She decides not to do so at ***t2***, and her decision is preceded by some appropriate sign Z, occurring at ***t0***, that makes it probable that she will not insult Stewart (perhaps a relaxation of certain facial muscles). Crispin detects Z and, understanding its significance, does not change his opinion concerning Helen's character. However, he might have done so had he not seen it.[2]

Let 'p' stand for 'Crispin does not change his opinion concerning Helen's character', and 'q' for 'Helen decides not to insult Stewart'. In that case, Np is true because Helen cannot ensure that p is false (that Crispin changes his opinion about her). If Helen decided to insult Steward, the sign would not have occurred and Crispin *might* have changed his opinion concerning Helen's character. But it is not the case that he *would* have done so. Likewise N(p → q) is true (if we take N to apply only to Helen's abilities) because in order to make the material implication false one must make the consequent false and the antecedent true and Helen cannot do that in this case. For, she can make q false by insulting Steward, but she cannot then ensure the truth of p because in that case Crispin might have changed his opinion concerning her character. According to Beta, we should conclude that Helen cannot insult Steward (Nq), but by assumption, she can do that. Therefore, this case seems to show that Beta is invalid for this interpretation of N, because it leads from the true premises to the false conclusion.

But, what is the explanation of Beta's failure in this case? Clearly, its failure is the consequence of the fact that although Helen does not have the power to ensure that p is false and does not have the power to ensure that (p→ q) is false, she has the power to ensure that

[2] Timothy O'Connor, *Persons and causes*, (Oxford: Oxford University Press, 2000), 8-9.

one of them is false. Thus, the validity of Beta requires not just that N applies to propositions in each premise, but also that it applies to their conjunction. For, it is sufficient for the falsity of the conclusion that one has a choice about the truth of one *or* the other premise, without having a choice about their truth separately. In other words, Beta is valid only if the following rule called Agglomeration is valid: Np, Nq ⊨ N(p ∧ q). But Agglomeration is not valid for N interpreted in the way suggested above, and so neither is Beta.

Thomas McKay and David Johnson have demonstrated the invalidity of Agglomeration in the following way. Suppose you have a fair coin and you did not toss it at a certain moment (although you could have). In that case it is true that 'the coin did not land heads at that moment and you had no choice about that' (Np) and it is true that 'the coin did not land tails at that moment and you had a choice about that' (Nq). (N here refers to the unavoidability of a particular state of affairs for a particular person at a particular time). However, it is clearly false that no one had a choice at that moment about the truth of the statement "the coin did not land tails and did not land heads," that is, of N(p ∧ q), because you had it in your power to ensure its falsity simply by tossing the coin.[3]

McKay and Johnson also demonstrated the invalidity of Beta by using the coin-toss example. This is my reconstruction of their demonstration. Suppose again that you were holding a fair coin at some moment, and decided not to toss it. If p stands for 'the coin did not land heads' and q for 'the coin was not tossed,' Np and N(p→q) are true because no one has or ever had a choice about the fact that the coin did not land heads, and no one has and or ever had a choice about the fact that the coin did not land heads andthat it was tossed is true (i.e. that p and not q is true). However, it is clear that Nq is true, because you had a choice about tossing the coin. Therefore, Beta again turns out to be invalid.

[3] My presentation of McKay and Johnson' counterexamples to Agglomeration and Beta are based on the presentation of their counterexamples by Warfield and Crisp. See Thomas M. Crisp and Ted A. Warfield, "The Irrelevance of Indeterministic Counterexamples to Principle Beta," *Philosophy and Phenomenological Research*, 61 (Jul., 2000): 177-179.

The question is, however, if Beta can be fixed so that it avoids these problems. Several strategies for doing that have emerged. Some authors such as Widerker, Warfield and Finch have suggested that Beta should be replaced with the following principle: (Np, $\Box(p \rightarrow q)$, $\nvdash$ Nq). This principle is immune to the above presented counterexamples since it is impossible to have the power to make a conjunction false, but lack this power with respect to one of the premises if the other premise is necessarily true. In other words, if one can ensure the falsity of a conjunction of a necessarily true proposition with a proposition which is not necessarily true, then one can ensure the falsity of the proposition which is not necessarily true. This seems like a natural solution to the problem because this new version of Beta does not entail Agglomeration.

However, this suggestion is not completely satisfactory. For the incompatibilist who accepts this principle must still rely on Agglomeration to get from NPo and NL to N(Po $\wedge$ L). The incompatibilist who accepts this reformulation of Beta can reply that this is not a problem because N(Po $\wedge$ L) is obviously true. However, as O'Connor observes, this answer is still not fully satisfactory because it is puzzling that Agglomeration "breaks down" in some cases for our ordinary notion of 'unavoidability' (the term O'Connor uses for the operator N).[4] In other words, it would be better if we could find an explanation of the failure of Agglomerativity and use it to resolve the problem.[5]

The reason why Agglomerativity fails in some cases is actually very simple. But, it is not clear if the problem can be solved in a satisfactory way. The main reason why Agglomerativity fails in some cases is that in order to possess one ability it is often necessary to have some other ability or set of abilities, but the standards for ascription of those abilities

[4] See O'Connor, *Persons and causes*, 13.

[5] There are two other potential problems for this version of Beta. One is that some philosophers think that N (Po $\wedge$ L) is not obviously valid (David Lewis, for instance makes that claim). The other problem is that this version of Beta looks very similar to some invalid principles for epistemic necessity. For comparison of these principles with Beta see Michael Slote, "Selective Necessity and the Free Will Problem," *Journal of Philosophy* 79 (Jan., 1982): 5-24.

are not always equally strict. An example of this phenomenon is the relation between our ability to control our bodily movements and to control processes within our bodies. The former ability requires the latter. For instance, to raise my arm I must cause the occurrence of certain process in my nerves and in my muscles. However, although I am able to raise my arm intentionally I am not able to *intentionally* cause those processes because I know nothing about them. Similarly, the ability to render false a conjunction requires the ability to render false one of its conjuncts, but the standards for ascription of the former ability may not be as rigorous as the standards for ascription of the latter ability. The coin toss example illustrates this claim. To be able to *ensure* the falsity of the conjunction "the coin does not land heads and the coin does not land tails" one needs to be able to render false at least one of the conjuncts, but not to *ensure* that either of them is false. This phenomenon makes the logic of unavoidability "unstable." It creates circumstances in which the failure of Agglomerativity or "slippage," (as O'Connor calls it) is possible.

An obvious strategy for avoiding this problem consists in weakening of the notion of ability on which the notion of unavoidability is 'parasitic.' For, the weaker the notion of ability is, the less space there is for the slippage. In addition, the weakening of the notion of ability does not necessarily create problems for the Consequence Argument because if the argument shows that under determinism no one can do otherwise in a weaker sense, it also shows that no one can do otherwise in the stronger sense. The question is only if there is an ordinary sense of ability which is sufficiently weak to avoid slippage and if not whether there is some non-standard notion of ability which preserves the truth of the premises of the Consequence Argument. I explore the 'weakening strategy' in the following section.

1.1.1.2 Weakening of the Notion of Ability

Most incompatibilist who pursue this strategy suggest a maximally weak notion of ability. Thus, O'Connor suggests what he calls the "ability in the minimal sense". According to O'Connor, "one is able to make it the case that either *p* or not-p in this sense just in case it is open to one so to act (reliably or not) that it *might* be the case that *p,*and open to one so to act that it *might* be the case that not-p."[6] Similar definitions have been offered by van Inwagen, Fischer and Ginet. This definition of ability eliminates the problems with Agglomeration. For, if it is open to me so to act that it might be the case that not (p ∧ q), it must be open to me so to act that it might be the case that p or that it might be the case that q. In other words, if my action is consistent with the fact that not (p ∧ q), it must be consistent with not p or not q.

However, the problem with this solution is that on this interpretation of ability the premises (concerning the laws of nature and the past) of the Consequence Argument are no longer noncontroversial. For, they say that it is not open to anyone to act (if one cannot act) in such a way that the proposition expressing the laws of nature or about the past might not be true. In other words, on this interpretation of ability, premises simply state that the only courses of action that a person is able to pursue are those consistent with the actual past and the actual laws of nature. However, this is exactly what compatibilists deny. In their view, to be able to do otherwise (in the *actual* circumstances), one's doing otherwise does not have to be consistent with the actual laws of nature and the past. For, according to compatibilists, the agent's ability to do otherwise (in the actual circumstances) is consistent with the truth of the conditional that if he acted otherwise some actual fact would not have been a fact. Compatibilists typically accept some sort of conditional analysis of ability to do otherwise according to which the agent would have done otherwise if he chose or wanted to do

[6] O'Connor, *Persons and Causes*, 30.

otherwise. And it is consistent with this analysis that the past has to be different for the agent to do otherwise.

Does this mean that those incompatibilists who accept the minimal notion of ability actually beg the question against compatibilists? In my view that is not the case and I will try to show it in the second part of this chapter. Nevertheless, I think that accepting the minimal notion of ability weakens the Consequence Argument considerably. For, in that case the argument is no more based on premises that anyone would accept. The incompatibilist who argues for his position in this way has a burden of showing that his premises are true.

So, let us see if there is some other notion of ability weaker than the ability to ensure but stronger than the minimal notion on which Beta is valid and the premises of the Consequence Argument uncontroversial. The concept of ability which is slightly weaker than the ability to ensure that something will happen is the concept of ability to unintentionally cause something to happen. Could this notion of ability help to eliminate the problems with Agglomeration? The answer seems to be no. For, this suggestion does not eliminate trouble with indeterministic counterexamples to Beta and Agglomeration. In the coin toss case, for instance, (assuming that the coin tossing process is genuinely indeterministic) we can cause conjunction to be false, but we cannot cause either conjunct to be false, either intentionally or unintentionally.

But, consider even weaker concept of ability: the concept of ability to raise the probability of a particular outcome or to causally contribute to a particular outcome. On this understanding of ability the coin toss case is not a counterexample to Agglomeration. For, to be able to contribute to the falsity of the relevant conjunction, one must at least be able to contribute to the falsity of a particular conjunct. Unfortunately, Kadri Vihvelin has shown that this notion of ability is also inadequate with the following counterexample to Beta:

> The lottery is taking place and you have no way to influence the outcome of the drawing and whether it will take place. The only thing you can do is to buy a lottery ticket. In that case, it is true that you have no choice about whether the number on your ticket will be drawn, and you don't have a choice about whether it is true that if your ticket is not drawn, you will not win the lottery. However, you do have a choice in the sense explicated above whether you will win the lottery in this sense, because you can contribute to your winning by buying the lottery ticket (even if your number has not been drawn).7

Now, since the only notion of ability that is weaker than this one is the minimal notion, it is clear why incompatibilists who pursue the weakening strategy base the argument on that notion. In other words, if we just focus on the interpretation of the operator N and the corresponding notion of ability in the attempt to avoid problems with Beta, we must accept the minimal notion of ability.

However, there is another, in my view, more promising strategy that incompatibilists can pursue. It consists in restricting the scope of Beta to deterministic context.

1.1.1.3 Restricting the Scope of Beta and Agglomeration

As we have seen, most counterexamples to Beta and Agglomeration are situated in indeterministic settings. For that reason, in a paper in which they argue for the irrelevance of indeterministic counterexamples to these modal principles, Warfield and Crisp have suggested that the scope of Beta and Agglomeration should be restricted to deterministic

[7]Kadri Vihvelin, "The Modal Argument for Incompatibilism," *Philosophical Studies* 53 (1988): 239.
Perhaps this counterexample could be neutralized by arguing that it is not true in this case that you have the ability to causally contribute to winning the lottery because you don't have the opportunity to do that. After all this would not be a totally ad hoc solution because all participants in the debate agree that ability which is in question in the Consequence Argument is the ability which includes the opportunity to perform the action. The problem with this suggestion, however, is that it seems to lead to a fatalistic understanding of abilities. Namely, it would turn out that one is able to contribute to coin's landing heads only if the coin does lands heads. In other words, this kind of defense of Beta makes sense only if we assume that only that which is the case can be the case.

scenarios.[8] In other words, they have suggested that Beta should be replaced with the principle Delta:

Delta: D, Np, N (p→q) ⊨ Nq

Obviously, Delta is immune to the counterexamples involving indeterministic processes. However, according to Eric Carlson, that is not the case. In his view, this revision of Beta does not fully eliminate the problem with the coin toss example not just because one cannot *ensure* that the coin will fall heads or that it will fall tails(or do it intentionally) even if the setting is deterministic, but because the truth value of the relevant counterfactuals is indeterminate. First, he says that if the agent had tossed the coin, it is not determinate whether the coin would have fallen head or tails because it is not determinate what the laws of nature and the past would be in that case. Second, he says that if the agent tossed the coin, the laws of nature might have been indeterministic. Thus, even in a deterministic world it might not be true that if the agent tossed the coin he would cause it to land heads or he would cause it to land tails. Therefore, we cannot say that if under determinism the agent has the ability to (unintentionally) toss a coin in such a way that it falls heads if or in a way that it falls tails.[9]

I must admit that I don't know enough about counterfactuals to give a direct answer to Carlson's objections, (although his objections seem quite suspicious). However, I think I have a good indirect reply to his objections. For, if Carlson is right about the nature of counterfactuals, Vihvelin's lottery example fails to show that Beta is invalid on the reading of ability as the power to causally contribute to the occurrence of a certain outcome. To see this notice that in her counterexample the truth of the first premise depends on the truth of the

[8] See Thomas M. Crisp and Ted A. Warfield, "The Irrelevance of Indeterministic Counterexamples to Principle Beta," *Philosophy and Phenomenological Research* 61, No 1 (Jul., 2000): 173-184.
[9] See Eric Carlson, "Counterexamples to Principle Beta: A Response to Crisp and Warfield," *Philosophy and Phenomenological Research* 66, No. 3 (May, 2003): 734-736.

counterfactual (or rather 'semifactual') that even if the person bought the lottery ticket, it would have no influence on the outcome of the drawing of the winning number. For, the drawing process was causally isolated from her buying of the lottery ticket. However, if we accept Carlson's first or second argument we cannot exclude the possibility that if the person bought the ticket the laws of nature and the past would imply her number being drawn or that her buying the ticket would influence her number being drawn indeterministically by a chain of incredible coincidences. So, Carlson's success in refuting Warfield and Crisps defense of Beta seems to have a price that from the compatibilist's perspective may be too high.

Therefore, as far as I can see, incompatibilists don't need to accept the minimal notion of ability in order to defend the Consequence Argument. I see no decisive reason why the argument cannot be valid on some stronger conception of ability such as the ability to cause or causally contribute to an outcome on which premises of the argument are plausible even for the opponents of the argument. It is difficult, though, to show which sense of ability is exactly the one which can serve the incompatibilist's purpose. However, that does not seem to be a big problem as long as we have no reason to doubt that there is *some* notion of ability meaning (stronger than the minimal) that can do the work.

I turn now to discussion of van Inwagen's first argument which will show even more clearly the nature of the objection that the argument rests on some sort of equivocation.

1.2 Van Inwagen's 'Non-modal' Argument

Van Inwagen's non-modal argument or his first argument (as it is often referred to) is an argument about the power of a *particular* person to perform a *particular* action at a particular moment in a deterministic world. Van Inwagen describes the person and the action in the following story:

> Let us suppose that there was once a judge who had only to raise his right hand at a certain time, T, to prevent the execution of a sentence of death upon a certain criminal, such a hand-raising being the sign, according to the conventions of the judge's country, of a granting of special clemency. Let us further suppose that the judge—call him `J'—refrained from raising his hand at T, and that this inaction resulted in the criminal's being put to death. We may also suppose that J was unbound, uninjured, and free from any paralysis of the limbs; that he decided not to raise his hand at T only after a suitable period of calm, rational, and relevant deliberation; that he had not been subjected to any "pressure" to decide one way or the other about the criminal's death; that he was not under the influence of drugs, hypnosis, or anything of that sort; and, finally, that there was no element in his deliberations that would have been of any special interest to a student of abnormal psychology.[10]

According to van Inwagen, it is possible to show that in spite of what we would ordinarily think, the judge in this story could not have raised his hand at T if determinism is true. And since the judge does not differ in any relevant way from any agent we normally consider as being to perform such action, we can generalize this conclusion to all agents in deterministic worlds. According to van Inwagen, we can show that judge J was not able to raise his hand at T with the following argument (Po and L have the same meaning as in the modal argument):

(1) If determinism is true, then the conjunction of **Po** and **L** entails **P**.

(2) It is not possible that **J** have raised his hand at **T** and **P** be true.

(3) If (2) is true, then if **J** could have raised his hand at **T**,
J could have rendered **P** false.

(4) If **J** could have rendered **P** false, and if the conjunction of **Po** and **L** entails **P,** then **J** could have rendered the conjunction of **Po** and **L** false.

(5) If **J** could have rendered the conjunction of **Po** and **L** false, then **J** could have rendered **L** false.

[10] Van Inwagen, *An Essay on Free Will*, 69.

(6) **J** could not have rendered **L** false.

(7) If determinism is true, **J** could not have raised his hand at **T**.[11]

Unlike the previous argument, this argument does not rely on any controversial rules of inference because it uses only rules valid in first-order extensional logic. Thus, the argument must be valid. But is the argument sound? That is, are the premises of the argument true?

Everyone agrees that the first three premises are true. As in the previous argument, premise 1 is just a consequence of a widely accepted definition of determinism. Premise 2 is true because if J did not raise his hand at T, P would not express the actual state of the universe at T. Premise 3 is true because it seems clear that one can render a proposition false if one can change the state of affairs expressed by that proposition.[12]

However, according to critics, premises (4) – (6) are problematic. Just like in the case of the modal argument, they argue, problems come to surface when we begin to unpack the meaning of the expression "can render false." Again, the main problem is supposed to be that there is no single interpretation of this phrase on which all premises of the argument are true. Thus, they say that if the expression is interpreted in a way that we usually interpret claims about abilities, that is, if we interpret it as the claim about the ability to causally influence things or bring about something in some robust sense, either the premise (4) or the premise (5) is false. But if we interpret it in some technical sense, or minimal sense, on which these premises are obviously true, the premise (6) becomes false.

[11] Ibid. 70.

[12]Van Inwagen defines the phrase 's can render (proposition) *p* false' in the following way: "It is within s's power to arrange or modify the concrete objects that constitute his environment in some way such that it is not possible in the broadly logical sense that he arrange or modify those objects in that way and the past have been exactly as it in fact was and *p* be true." Van Inwagen, *An Essay on Free Will*, 68.

It may be sufficiently clear from the discussion of the modal argument above what I mean by ordinary, strong notion of ability and the ability in the technical or weak sense. However, for the sake of clarity I give here more precise definitions of these senses of ability to do something:

> **Strong ability**: ability with respect to an event such that if a person exercised that ability the person's action would either be that event or it would bring it about causally.
>
> **Weak Ability**: ability with respect to an event such that if a person exercised that ability, the event in question would occur (or would have occurred).[13]

As David Lewis points out, these two senses or types of ability correspond to two theses about our ability to influence the laws of nature: the weak and the strong thesis. The weak thesis says that we are able to do something such that a law of nature would not be a law of nature. The strong thesis says that we are able to violate the laws of nature (an analogous distinction can be made concerning the ability to influence the past before one was born, but I will talk about that later). According to Lewis, it would be crazy to say that someone has the ability to violate the laws of nature, but there is nothing problematic in saying that someone could do something such that the laws of nature would be different. In other words, on the strong reading of ability premise 6 is obviously true, but that is not so on the weak reading of ability. However, according to Lewis, the premise 5 of the argument is false on the strong reading of ability.

Why does Lewis think that the premise 5 of the argument is false on the strong reading of ability? He thinks so because he thinks that the fact that someone has the power (in

[13] See David Lewis, "Are We Free to Break the Laws," *Theoria* 47 (1981): 113-21.

whichever sense) to do something which entails that something else is the case, entails only that the person has the weak ability with respect to that other thing, and because the fact that one has the weak ability with respect to something does not entail that one has the strong ability with respect to that. Thus, the ability to do something such that a law of nature would be broken does not entail that if one exercised that ability one's action itself would be a law-breaking event or that it would cause some law-breaking event.[14]

What about the premise (4)? Why do some philosophers think that on the strong reading of the expression "can render false" this premise must be false? This premise is an instance of the general principle that if one can render some proposition false than one can also render false any logical antecedent of that proposition. Philosophers who object to premise (4) reject this principle. They argue that if this principle were true, in deterministic worlds we would not be able to do what we actually do which seems obviously false. The principle seems to have this implication because by doing what we do we are actually rendering many propositions false (all the propositions incompatible with what we do), whose logical antecedents are, if determinism is true, propositions about the past before we were born.[15] According to the principle, we are able to do what we do only if we are able to render false those logical antecedents. But, since we are not able to render them false, we are not able to do what we actually do.

In reply to Narveson, who first raised this objection, van Inwagen admits that his view has the strange consequence that we can render false some propositions about the past before we were born. Namely, his theory implies that we can render false all *false* propositions about the past. In fact, this consequence simply follows from his account of 'can render false.'

[14] I will explain later why Lewis thinks so.

[15] This sort of objection was first raised by Jan Narveson, See Jan Narveson, "Compatibilism Defended," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 32, No 1(Jul., 1997): 83-87.For a recent development of this objection see Ferenc Huoranszki, *Freedom of the Will: a Conditional Analysis* (New York: Routledge, 2011), 25-26.

However, he does not see this as a reason for abandoning the principle. He says that the principle is analytic. In his view, we should rather see that there is no problem in saying that we can render some false propositions about the past false. This is so, because false propositions do not play any role in his argument.[16]

However, in my view, van Inwagen's reply is not very convincing. For, it sounds totally weird to say that someone has such ability. Perhaps the idea is that rendering false an actually false proposition does not require much of ability. That is, the idea might be that having ability in that sense just means having the ability in the weak sense. However, the question is then why couldn't we say that in that sense we have the ability to render false some *true* propositions about the past? And how should the proponent of the argument respond to the above criticism if he insists on the strong reading of ability?

In my view, the answer is that he should reject the general principle referred to here and replace it with the following principle: If we can render false some proposition which is a logical consequence of some *true* proposition, then we can also render false the latter proposition. I think that this principle serves better the incompatibilist's purposes than the principle van Inwagen accepts not only because it does not entail the strange claim that we are able to render false all false propositions about the past, but also because it enables us to distinguish between two important questions: 1) the question of the relation of our ability to do what we actually do to its logical antecedents and 2) the question of the relation of our abilities to perform action we don't actually perform and factors which entail that we will not perform them. We can see the difference between these two questions by reflecting on the following two instances of these questions:

[16] Van Inwagen, *An Essay on Free Will*, 68.

1) Given that I was born with a certain defect, and given that not being born with that defect is a necessary condition for playing basketball, can I play basketball?
2) Since there is a gene which I have and I could not prevent having which is sufficient for my playing basketball on a certain day, do I have the power to play basketball on that day?

I think that these two questions are obviously different. What I want to know when asking the first question is whether I can perform an alternative action. It is a question about my possession of some ordinary power. On the other hand, what I want to know when asking the latter question, if the question makes sense at all, is whether I can be the right kind of source of my actions, whether I have ultimate control over what I do, and whether I can be ultimately responsible for what I do. However, only the principle that I have just presented enables us to focus on the former question.

Thus, I think we have not yet seen a good reason for rejecting premises (4) and (5) of van Inwagen's non-modal argument. These premises seem plausible on both the weak and on the strong reading of the expression 'can render false.' For that reason, I think that the objection that the argument must rest on equivocation fails.

However, I might be mistaken that the premises (4) and (5) are true on the strong reading of "ability". Perhaps there are counterexamples to the principles on which they are based which I have not considered yet. After all, how can I be sure that they are true on the strong reading of ability if many great philosophers who thought about this problem for a very long time seem to be willing to concede that these premises are true only on the weak reading of ability? For that reason, it is necessary to examine the compatibilists' objection that on this reading premise (6) of the argument is false. More precisely, it is necessary to

examine the compatibilist claim that given this reading of ability, the incompatibilist cannot say that the premise is true without begging the question because this premise is acceptable only to those who already accept incompatibilism.

1.2.1 Why the Consequence Argument Does not Beg the Question

What begging the question exactly means is not an easy thing to say. However, the following definition of begging question offered by John Martin Fischer sounds very plausible to me: an argument is question begging if and only if the only reason for accepting a premise of the argument is the fact that one accepts the argument's conclusion. I will not here attempt to justify this account because it seems to me intuitively very plausible. I just want to show that if this account of begging the question is correct, the Consequence Argument does not beg the question, even on the weak reading of ability, because we have reasons independent of incompatibilism to think that no one can have the ability in the weak sense to render false the laws of nature or the distant past.[17]

1.2.1.1 A problem with the Local Miracle Compatibilism

One obvious reason to think that no one could have such ability is that otherwise it would be difficult to show that we are not able to do some things that we obviously are not able to do. For instance, it might be difficult in that case to explain why I people cannot walk on water or why someone could not build a spaceship that would travel faster than light.[18] A

[17] However, this does not mean that the argument represents a refutation of compatibilism because the reasons for accepting the premises can only be reasons that make them plausible, rather than obviously true. If that is the case then even though the argument does not beg the question, it only shows that it is plausible to think that determinism and ability to do otherwise are incompatible. That is what I want to show in this chapter.
[18] This is van Inwagen's example. See van Inwagen, *An Essay on Free Will*, 62.

natural explanation of our lack of these abilities is that exercising them is incompatible with the laws of nature. That is, it is natural to believe that we cannot do those things because doing would imply that something that we consider to be a law of nature is not a law of nature. But, if we accept compatibilists suggestion that we have abilities the exercises of which implies that some actual law of nature is not a law of nature, on what grounds can we deny that we can make things that can move faster than light or do other miraculous things?

David Lewis has offered an ingenious answer to this question. In his view, things which we obviously cannot do are things which would themselves be, or would cause "law-breaking" events. An event is law-breaking, according to Lewis, if it is such that if it were to occur, *necessarily* some law of nature would be broken. Thus, if one were to create a machine that would cause particles to move faster than light, one would thereby cause a law-breaking event. Similarly, if one were to move one's hand faster than light that would itself be a law-breaking event. But, according to Lewis, we have no reason to think that the performance of some ordinary action that has not actually occurred would either be a law-breaking event or cause some law-breaking event in a deterministic world. Of course, if someone performed some action that he or she did not actually perform, given determinism, either the laws of nature or the past would have to have been different. Lewis thinks that the laws of nature would have been different. Thus, if a person acted otherwise, a divergence from the laws of the actual world or a miracle relative to the actual world would have happened. But what is important, according to Lewis, we have no reason to think that the agent's action would itself be that miracle. Moreover, according to Lewis, we have a reason to think that the miracle would have occurred prior to the occurrence of the agent's action. Let me clarify this.

That the divergence in laws of nature, or local miracle, would have occurred before the agent decided to perform an alternative action follows from Lewis' method of evaluating counterfactuals. According to Lewis, a counterfactual is true iff there is no world in which the

antecedent occurs and the consequent does not occur which is closer or more similar to the actual world than any world in which both the antecedent and the consequent occur. According to Lewis, this account together with plausible assumptions about the criteria for determining the similarity of worlds and facts about the relevant sets of worlds, yields the result that whenever we consider events that have not actually occurred (under the assumption of determinism) we must assume they were preceded by local or divergence miracles. That is so because, under determinism, for an event that has not actually occurred to have occurred, either the whole history up to the occurrence of that event would have to have had been different or a small miracle (breaking of a law of nature) allowing for a divergence from actuality would have to have had occurred.[19] Now since according to Lewis, a small divergence from the laws of nature (the small miracle) is much smaller a departure from actuality than divergence in the entire history up to the event in question, for something that hasn't actually happened to have happened, it would have had to be preceded by a small miracle ((if it is not an event which is itself law-breaking).

Therefore, the compatibilist who believes that we have the weak ability to render false the laws of nature seems to have the resources to explain the difference between things that we intuitively can do and things that we intuitively cannot do even if determinism is true. Actions that we cannot perform are actions that would themselves be law-breaking or would cause such events, while the actions we can perform are just the ordinary actions which would have been preceded by some small miracle if we decided to perform them. For obvious reasons, Lewis' theory is usually called the "local-miracle compatibilism. Thus, it seems that the distinction between weak and strong abilities is not just an ad hoc construct for saving compatibilism.

[19] This would, of course, be breaking of a law of nature only relative to the actual world. In the alternative world there would be no violation of laws nature.

However, according to Helen Beebee, although Lewis offered an account of why having a weak ability does not imply having a strong ability, he has not really explained why our ordinary abilities, in the framework of his theory, could not enable us to violate the laws of nature. According to Beebee, this is the case because it is not clear why an alternative action would have to be *preceded* by a small miracle in deterministic worlds. That is, in her view, we have no good reason to think that our actions themselves could not constitute "small miracles". As Beebee notices, our actions cannot be law-breaking (in themselves) in the sense Lewis gives to that expression, that is, they cannot be law-breaking in the sense that they "wear their miraculous nature on their sleeve." However, as Beebee points out, they could be law-breaking in the sense in which a divergence miracle is law-breaking: they can be law-breaking "in the circumstances." Therefore, "local miracle compatibilism," has an unpalatable consequence: it seems to leave space for the power to break the laws of nature.[20]

An interesting reply to this criticism has been presented by Peter A. Graham. In his reply, Graham distinguishes between two versions of local miracle compatibilism: the modest local miracle compatibilism which he identifies with Lewis's view, and more ambitious local miracle compatibilism. According to Graham, a modest local miracle compatibilist, like Lewis, does not claim that all of our ordinary abilities or abilities that we normally take ourselves to have are compatible with determinism, but only that such abilities can be compatible with determinism. The more ambitious local miracle compatibilist believes that determinism is not a threat to any of our ordinary abilities. As Graham describes them, both types of local miracle compatibilist agree with Beebee that an agent's (ordinary) action might be a divergence miracle in some situations. But they block the inference to the conclusion that according to local miracle compatibilism we have powers to perform miracles in a

[20] See Helen Beebee, "Local Miracle Compatibilism." *Noûs* 37 (2003): 258-277.

different way. The modest local miracle compatibilist denies that an agent would have the ability to perform an (ordinary) action if its performance would constitute a local miracle in the circumstances. On the other hand, the more ambitious local miracle compatibilist believes that the claim that one is able to do something and the conditional that if one were to do it one would perform a miracle, do not jointly imply that one has the ability to perform a miracle.

Graham defends the more ambitious view by pointing to the fact that the ability claims and counterfactual are evaluated by looking at different possible worlds. When we evaluate ability claims we consider whether there is a possible world in which we perform the action which is accessible to us from the actual world. On the other hand, when we evaluate counterfactuals, we consider worlds that are most similar to the actual world. More precisely, we consider whether the closest possible worlds in which the antecedent is true are also worlds in which the consequent is true. However, as Graham points out, the worlds in virtue of which the counterfactual is true may not be accessible to us even though some worlds in which the antecedent of the counterfactual is true are accessible to us. He offers several examples in which this seems to be the case. Thereby he shows that a claim that a person is able to do something and the counterfactual that if the person performed the action in question the person would break the laws of nature, do not jointly entail that the person has the power to break the law of nature.

Graham's defense of the less ambitious view consists in his defense of the right of the compatibilist simply to deny (without some special explanation) that the agents would have certain ordinary abilities in situations in which exercising them would be law-breaking (a local miracle). According to Graham, the incompatibilist cannot complain that the compatibilist does not have an explanation for why one would lack the ability to do something that would in the circumstances amount to breaking of a law of nature, because the incompatibilist also has nothing to say about why people lack that power. For, according to

Graham, the incompatibilist takes it as "a platitude" that no one can break the law of nature. Besides, according to Graham, the incompatibilist cannot complain that the modest local miracle compatibilist makes the possession of ordinary powers dependent on some counterfactual circumstances because the incompatibilist also thinks that our ordinary powers depend on the physical facts (on whether what we do is determined by laws and the past).[21]

Does this help eliminate the worry Beebee has raised about the local miracle compatibilism? In my view, the answer is no. The problem with Graham's defense of the ambitious local miracle compatibilism is that he does not show that the world in which it is true that if the action occurred it would be the small miracle, cannot be the same as the world which makes the ability claim true. And that is precisely what worries Beebee. Besides, one might also worry about the fact that some such counterfactual is true in some world.

The problem with Graham's defense of the modest version of the local miracle compatibilism is less obvious. Graham's observation that the incompatibilist cannot complain that the local miracle compatibilist takes our ordinary abilities to depend on counterfactual circumstances because he also takes those abilities to depend on some physical facts seems correct. However, it is not correct, as Graham claims, that the incompatibilist takes it as a platitude that no one can *violate* the laws of nature. The incompatibilist has an explanation for that fact. The explanation is that the laws of nature *constrain* the agent's abilities. Of course, the incompatibilist does not say anything about the way in which the laws of nature constrain abilities. However, that is a different matter. The important point is that the compatibilist has *nothing* to say to support the claim that no one can break the laws of nature because he does not think that those laws constrain the agents' abilities.

[21] See Peter, A. Graham, "A Defense of Local Miracle Compatibilism." *Philosohical Studies: An International Journal for Philosophy in the Analytic Tradition* 140 (2008): 65-82.

Beebee explains nicely why this problem occurs for the local miracle compatiblist by pointing to two conceptions of the law of nature. According to Beebee, the deeper explanation of this result is that there is no understanding of the notion of the law of nature which can make sense of the distinction between the weak and the strong senses of ability to render the proposition expressing the law of nature false. Thus, if we understand the law of nature in the Humean sense, as some sort of generalization about what actually happens in every instance of the world's existence, it is clear why we could have the ability in the weak sense but not why we would have to lack the ability in the strong sense. On the other hand, if we understand the laws of nature in the necessitarian way, it is clear that we could not have abilities in the strong sense, but not how we could have the ability in the weak sense. For on the necessatarian view of laws

Beebee's point becomes even clearer if we think about the problem of explaining the harmony between the agents' choices and factors which determine those choices. There are two natural explanations of this harmony: the Humean and the Necessitarian. The former is that agents necessarily do what the laws of nature proscribe because the laws of nature depend on their actions and not vice versa. The latter answer is that they necessarily do what the laws say they will do because they cannot do otherwise, that is, because the laws constrain the powers of everything including the powers of agents. Neither of these two conceptions of the laws of nature seem friendly to compatibilism (considered here). Compatibilism seems to require some concept of the natural law which would make sense of the idea that there is some kind of mutual dependence between the agents and the laws. However, the question is whether such a conception is possible.

This I think supports the claim that there are independent reasons (or a reason) for thinking that the premise of the argument which says that no one has the power to do something that requires the falsity of some law of nature is true. I think that the same is true

about the alternative premise which says that no one has the weak ability to render false the past. I explain why I think so in the following section.

1.2.1.2 The Problem with the Different Past Compatibilism

As we have seen some philosophers, most notably David Lewis, believe that acting differently in deterministic worlds would require that some law of nature is not a law of nature. More precisely, they think that if one were to do otherwise a small divergence in the laws of nature would have occurred just prior to the performance of that action. But, some philosophers think about counterfactual situations of this sort differently. They think that if one acted differently in a deterministic world, the entire past rather than some law of nature would have been different. For incompatibilists, the truth of this claim is equally good reason to believe that ability to do otherwise is incompatible with determinism as it is the truth of the claim that if one acted differently some actual law of nature would not be a law of nature. However, some compatibilists do not think that the truth of the "different past" or "backtracking" counterfactual entails the lack of the corresponding ability to do otherwise. Andre Gallois states this view in the following way:

> We cannot argue that simply because a proposition expresses a state that the world was in prior to an individual's birth, that individual could not have rendered that proposition false, if all that is meant by having the capacity to render a proposition false is having the capacity to perform an action whose performance would be sufficient to insure its falsity.[22]

[22] Andre Gallois, "Van Inwagen on Free Will and Determinism," *Philosophical Studies: An International Journal for Philosophy in Analytic Tradition,* Vol. 32, No. 1 (Jul., 1977): 103

Gallois supports this view with an argument very similar to the one earlier considered for the falsity of the premise (4) of the non-modal version of the Consequence Argument. Here is Gallois' presentation of the argument in question:

> Suppose that J, the judge in van Inwagen'sexample, had raised his hand at Time T. Then, since the conjunction of Po with a set of natural laws entails that J refrained from raising his hand at T, Po would be false. Moreover, if J had raised his hand at T, then, given the truth of determinism, some proposition (let us call it Po) would have been true, where Po in conjunction with L entails that J raised his hand at T and Po expresses a state of the world at a time prior to J's birth. Consequently, in refraining from raising his hand at T, J rendered Po false. That is, J could have and in fact did refrain from performing an action, where refraining from performing that action was a sufficient condition for Po being false. So we cannot argue that simply because a proposition expresses a state that the world was in prior to an individual's birth, that individual could not have rendered that proposition false, if all that is meant by having the capacity to render a proposition false is having the capacity to perform an action whose performance would be sufficient to insure its falsity.[23]

The key element in this argument (as in Narveson's argument against the premise (4) of van Inwagen's non-modal argument presented earlier) is the observation that by doing what we actually do we render certain propositions about the present state of the world false and the falsity of those propositions implies the falsity of propositions about the past before we were born (whose truth implies the truth of the propositions about the present). Narveson used this observation to show that it cannot be true that our *strong* ability to render a proposition false transfers to the antecedents (sufficient conditions) of that proposition. Gallois uses the above mentioned observation to establish the conclusion that we obviously have abilities to render some (false) propositions about the past false in the *weak* sense and on that basis challenges the claim that we don't have the weak ability to render false true propositions about the past.

[23] Ibid.

In response to Narveson's argument I said that we need to replace the general principle that if one can render Q false and P implies Q, then one can render P false, with the principle that if one can render Q false and P, which is true, implies Q, then one can render P false. Only the latter principle, I argued, captures the idea that when a necessary condition for performing an action is absent, we can perform that action only if we can bring about that condition.

My response to Gallois is that the weak ability to render false a false proposition about the past and the weak ability to render false a true proposition about the past are not the same abilities. The former ability is the ability the exercise of which requires satisfaction of some condition that is already satisfied. This is the case with the ability to render false a proposition about the present which is implied by a proposition about the past which is false or with the ability to render true a proposition about the present the antecedent of which is a true proposition about the past. In my view, this kind of ability is intuitively different from the ability to render false a (true) proposition about the present which is implied by a true proposition about the past. The first is the ability to do something when we have everything that we need for the performance of the action, while the latter is the ability to do something when something that is necessary for the performance of the action is absent.

Now, Gallois explains the persuasiveness of the claim that no one can or could have render false the proposition about the past by pointing to the fact "that 'could have in its normal usage is linked to the appropriateness of deliberation."[24] We think that that we cannot do anything about the past because it makes no sense to deliberate about the past. And it makes no sense to deliberate about the past because we cannot influence the past by our choices. It is not so with our future actions. For it makes sense to think that even if our actions are causally determined by the state of the world at some moment before we were

[24] Ibid. 103.

born, the causal chain leading to our actions includes our process of deliberation as a key element. Therefore, Gallois thinks that paying attention to the deliberative perspective can help us understand why the Consequence Argument seems persuasive and why it ultimately fails.

However, focusing on deliberation shows that the idea that we might have the weak ability to render false the propositions about the past has some very unpalatable consequences. For, as John Martin Fischer and Garrett Pendergraft point out in a recent paper, if we accept the idea that we have this ability, then it seems to follow that sometimes we have good reasons to do things which are obviously irrational. To show this they point to the fact that there are some backtracking conditionals which seem true. They offer the following example of such conditional that Fischer presented in his earlier work:

> Consider the example of the Icy Patch. Sam saw a boy slip and fall on an icy patch on Sam's sidewalk on Monday. The boy was seriously injured, and this disturbed Sam deeply. On Tuesday, Sam must decide whether to go ice skating. Suppose that Sam's character is such that if he were to decide to goice-skating at noon on Tuesday, then the boy would not have slipped and hurt himself on Monday.[25]

As Fischer and Pendergraft point out, if we assume that Sam can decide on Tuesday to go ice-skating, it seems that Sam has a reason to go ice-skating on Tuesday. Moreover it seems that Sam *ought* to go ice skating given that if he were to do that "the boy would not have slipped and hurt himself on Monday." However, deciding to do that would be clearly irrational given that he knows "that the accident did in fact take place on Monday."[26]Or as

[25] John Martin Fischer and Garret Pendergraft, "Does the Consequence Argument Beg the Question?" *Philosophical Studies* 166 (2013): 587.
[26] Ibid.

Fischer and Pendergraft notice, "to do so would seem to exemplify something akin to wishful thinking."[27]

The incompatibilist seems to be better positioned than the compatibilist to resolve this puzzling situation. The incompatibilist can grant that it could be the case that Sam has a power to decide to go ice skating on Tuesday and that it is true that if he were to do that the accident would not have occurred on Monday, but legitimately deny that Sam has a reason to decide to go ice-skating. This is so because he denies that worlds in which the backtracking counterfactual is true are among the worlds in which Sam goes ice-skating which are accessible to him. And he thinks so because he believes that we can make some event happen now only if we can "make the world contain everything that has happened before now plus that event after now."[28] On the other hand, since the compatibiist thinks that the world with a different past is accessible to the agent, it is not clear how the compatibilist can deny that Sam does not have a reason to decide to go ice-skating on Tuesday.[29]

Perhaps there is some way for the compatibilist to avoid this problem. The compatibilist could deny that the backtracking counterfactual is true. Perhaps Sam would act out of character if he was to decide to go ice-skating and it would not be the case that the accident had not taken place earlier. But as Fischer and Pendergraft point out, the story in question "can be filled in so that it is at least plausible that the backtracker is indeed true."[30] And as they point out this is enough for their argument.

[27] Ibid.
[28] Fischer and Pendergraft attribute this principle to Carl Ginet. Fischer and Pendergraft, "Does the Consequence Argument Beg the Question?" 588.
[29] At first sight, it seems that compatibilists and incompatibilists are in the same position as long as they agree that both the can claim and the backtracking counterfactual are true. But, the 'trick' is, so to say, that when the incompatibilists evaluates can claims he has to keep the past and the laws of nature fixed. In other words, the incompatibilist believes that his abilities are abilities to add to the given past in accordance with the laws of nature. So, it is clear that the truth of the counterfactual cannot be a reason for the incompatibilist to act irrationally.
[30] Ibid. 587.

Thus, it seems that there is a reason independent of the conclusion of the Consequence Argument to accept the premise that no one has the weak ability to render false propositions about the past state of the world before one was born. Therefore, even if the phrase "can render false" is understood in the weak sense, the Consequence Argument represents a threat to compatibilism.[31] The Consequence Argument does not show that it is impossible that compatibilism is true, but it shows that it is very plausible to think that compatibilism is false.

1.3 Conclusion

We can see now more clearly why the Consequence Argument has the reputation of the strongest argument for incompatibilism. Obviously, this is so because it is difficult to see where it goes wrong. In particular, the main objection according to which there is no interpretation of free will on which its premises are true and the inference rules valid does not hold. For, there are at least two versions of the argument - van Inwagen's modal and modal version – which are immune to this objection. Van Inwagen's modal argument seems to be immune to this objection because its premises are obviously true on the strong reading of 'ability,' and there are versions of its inference rules which are immune to counterexamples designed to show that the argument is invalid on the strong reading of ability. On the other hand, his non-modal argument seems to be sound because it uses non-controversial rules of

[31]On closer examination, I think we can see that this reason is very similar to the reason for rejecting the weak ability to render propositions about the laws of nature false. The problem in both cases is that it is difficult to distinguish between the weak and the strong ability in question. In the latter case this difficulty is manifested in the fact that from the deliberative perspective it does not make much difference when we say that we can do something such that our action would make it the case that the past is different or that we have a power to do something such that if we were to do it the past would be different because in both cases we seem to have a reason to exercise that power, and that is what is strange.

inference and it rests on very plausible premises which we have reason to accept independently of its conclusion.

Therefore, it is plausible to conclude that the ability to do otherwise and determinism are incompatible. Consequently, free will is possible if determinism is true only if free will does not involve ability to do otherwise or if it is possible under assumption that determinism is false. In the chapters that follow I will explore these alternative options.

CHAPTER 2: MORAL RESPONSIBILITY AND ALTERNATIVE POSSIBILITIES

Until recently, the claim that free will and moral responsibility require ability to do otherwise enjoyed the status of an axiom in moral philosophy. In other words, all philosophers agreed that the following principle is true:

> Principle of Alternative Possibilities (PAP): An agent is morally responsible for what she has done or omitted only if she could have done otherwise.

The conviction that this principle is a fundamental truth about free will was no doubt the main reason for the worry that we don't have free will and are not morally responsible if determinism is true. That was also the reason why the free will debate was focused on the question about the compatibility of ability to do otherwise and determinism. The focus of the debate changed in 1969 when Harry Frankfurt published his famous paper "Alternate Possibilities and Moral Responsibility."[32] In that paper Frankfurt challenged the idea that free will requires ability to do otherwise by examining the role PAP has in our practices of holding people morally responsible and by offering a counterexample to this principle. In another paper Frankfurt suggested an account of moral responsibility which does not refer to alternative possibilities.[33] Frankfurt's ideas, and in particular his counterexample, had a profound influence on the debate about free will and moral responsibility. They convinced many philosophers, both compatibilists and incompatibilists, that moral responsibility is essentially a matter of what goes on in the actual sequence of events leading to action. It

[32] Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66 (1969): 829-39
[33] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68 (1971): 5-20

motivated compatibilists to develop varieties of what has come to be known as 'semi-compatbilism' – the view that free will relevant for moral responsibility is compatible with determinism even if ability to do otherwise is not. On the other hand, it motivated incompatibilists to develop varieties of the so called 'source incompatibilism' - the view that determinism is not incompatible with free will because it eliminates alternative possibilities, but because it prevents one from being the right kind of source (the ultimate source) of one's action.

However, Frankfurt's arguments did not convince everyone. Those who were not convinced offered good reasons against the claim that his observations and his counterexample show the falsity of PAP. In response, Frankfurt's followers offered new examples which then gave rise to new objections. As a result of this exchange a very vigorous debate has developed and the question whether 'Frankfurt-style strategy' can show the falsity of PAP became and still is one of the main questions in the free will debate.

In this chapter I will argue that Frankfurt's argument against PAP fails. This result is important for my overall thesis – the thesis that we are sometimes directly free and responsible when we perform good actions for the right reasons but never when we perform wrong actions or act for wrong reasons. For, my argument for this thesis depends on the thesis that free will is asymmetric in the sense that it requires ability to do otherwise when we perform wrong actions or act for wrong reasons but not when we perform right actions for the right reasons,[34] which must be false if Frankfurt's argument is sound since its aim is to show that the ability to do otherwise is never required for free will and moral responsibility.

My defense of PAP from Frankfurt's attacks will consist in showing that his counterexample and modified versions of his counterexample to PAP fail. I will mostly rely on objections raised by other philosophers in trying to achieve this aim. But I will also add a

[34] In chapter five I will present an argument for the claim that we can make sense of the conjunction of this asymmetry thesis and incompatibilism by accepting the main thesis of my dissertation.

few observations of my own that will, I believe, strengthen the objections that other philosophers have raised.

I plan to achieve my aim in the following way. In section 2.2 I will present Frankfurt's argument. In 2.3 I will consider a problem for his argument which, in my view, helps to understand better his argument. Then in 2.4 I consider a reply to his argument by compatibilists who think that the ability to do otherwise is necessary for moral responsibility which helps to clarify the notion of ability to do otherwise is at stake in the argument. In section, 2.5, I will further clarify this notion by presenting an early reply to Frankfurt's argument. Then, in section 2.6, I will present what I consider to be a decisive objection to Frankfurt-strategy. In the rest of the chapter, that is, in the last four sections, I will consider attempts on behalf of Frankfurt's followers to meet this objection and argue that they all fail.

2.1 Frankfurt's Challenge to PAP

It is not difficult to see why PAP is so attractive. PAP has a great explanatory power. It explains our moral judgments toward people in a wide variety of cases. Here are a few examples. Imagine a soccer player who has acted violently toward a referee, say, by punching him in the nose after receiving a red card. At first sight, such a player is guilty and deserves blame. But, imagine that just before his wrongful act he suffered a nervous breakdown or that he was hypnotized and instructed by a hypnotist to perform that act. Under such circumstances it would be inappropriate to blame the player. Or imagine that a friend of yours who was supposed to return you the money you desperately needed, instead of doing that went to a casino and wasted it. Normally you would be angry and hold your friend responsible for such an act. But if he was a pathological gambler such attitude would not be appropriate.

The cases of a football player and of the gambler just mentioned are the paradigm cases of lack of responsibility. For, extreme distress, mind control or irresistible desires are generally considered to be factors that deprive agents of moral responsibility. These factors undermine agents' responsibility because they compel agents to do what they do and thus take away their free will. But why do we think that compelled agents lack free will? A natural answer to this question is that they lack ability to do otherwise. On the other hand, when such factors are not at work, that is, when people act freely, they always seem to have the ability to do otherwise.

However, Harry Frankfurt noticed that lack of ability to do otherwise is not the only common feature of typical cases in which the agents lack moral responsibility. The feature that philosophers before Frankfurt had failed to notice is that the factors which deprive agents of responsibility in those cases besides depriving them of ability to do otherwise also account for what they actually do. In other words, Frankfurt noticed that usually agents who lack free will not only lack alternative possibilities but also act as they do *because* they lack alternative possibilities. This discovery opened up space for the possibility that victims of compulsion or coercion lack free will not because they lack alternative possibilities but because their lack of alternatives possibilities explains what they do. That is, it opened doors to the possibility that alternative possibilities are not *per se* relevant for moral responsibility. Frankfurt believed that this is indeed so and in support of that claim presented the following example:

> 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 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. [35]

According to Frankfurt, this example shows that alternative possibilities are not necessary for moral responsibility. For, in his view, we can assume that Jones is morally responsible even though he cannot do other then what he actually does. We can assume that Jones cannot do otherwise, Frankfurt argues, because the example is flexible enough that anyone with a theory about the ability to do otherwise can add details to the examples that would make clear that according to that theory Jones is not able to do otherwise. For instance, one could suppose that Black would force Jones to do what he wants him to do by pronouncing a terrible threat, or by generating an "irresistible inner compulsion" in him by giving him a potion, or by putting him under hypnosis or even by direct manipulation of his brain. However, according to Frankfurt, the fact that Black would do that has no bearing on Jones's responsibility for his action. For, Black does not actually interfere with Jones's process of deliberation. He only lurks in the background ready to intervene if that turns out to be necessary.

Frankfurt explains why Black's inactive presence and his readiness to intervene have no bearing on Jones's moral responsibility in the following passage:

> 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 did it. It would be quite unreasonable to excuse Jones for his action, or to withhold the praise to which it would normally entitle him, on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did. He would have acted the same even if it had not been a fact. Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it.[36]

[35] Frankfurt, "Alternate Possibilities and Moral Responsibility," 21.
[36] Ibid. 22.

This passage expresses an important idea that a factor has no relevance for the agent's responsibility for an action unless it explains *why* the agent performed that action. For, if the agent did not perform or omit the action *because* of that factor, the factor cannot serve as an excuse for what he did and cannot be the basis for withholding blame or praise for what he did. This is so, Frankfurt argues in this passage, because he would do the same even if the factor in question had not been present.

Therefore, the reason way Frankfurt thinks that we can assume that Jones is morally responsible although he could not have done otherwise is that it is not possible for Jones to do otherwise in Black's presence, and because Black's presence does not explain his action. We can assume that Jones is morally responsible because the only difference between an ordinary situation in which we would hold the agent responsible and this case is the presence of Black which does not explain Jones's action. On the other hand, we can assume that Jones cannot do otherwise because we can assume that his doing otherwise is impossible.

Frankfurt's argument can be summarized in the following way:

1) In the absence of Black Jones is morally responsible for what he does.

2) Black's (mere) presence and readiness to intervene renders Jones unable to do otherwise.

3) A factor has no bearing on the agent's responsibility for an action unless it explains *why* the agent performed it.

4) Black's presence does not explain why Jones acts as he does.

Therefore,

5) Black's presence renders Jones unable to do otherwise, and Jones is morally responsible for his action.

Thus,

6) An agent can be morally responsible for his action even though he lacks ability to do otherwise.

Therefore,

7) Ability to do otherwise or the existence of alternative possibilities is not necessary for moral responsibility, that is, PAP is false.

If this argument is sound, compatibilists seem to be in a much better position than they were previously. For if free will does not require ability to do otherwise one of the main problems for compatibilism, the problem of compatibility of determinism and ability to do otherwise, is not a problem for them anymore. However, as Frankfurt notices, if this argument is sound, it does not automatically follow that free will and determinism are compatible. For if determinism is true there are factors that render actions inevitable (i.e. eliminate alternative possibilities) which *also* explain why agents perform them. Also, if this argument is sound we need an explanation of the lack of responsibility in typical cases, that is, in cases behavior which is the result of compulsion or coercion. Frankfurt explains that agents lack responsibility in these cases but not necessarily if their actions are causally determined because lack of ability to do otherwise deprives agents of responsibility if they performed act *only* because they could not do otherwise.

However, is Frankfurt's argument really sound? In what follows I will argue that it is not by showing that the premise 5 of his argument is false. That is, I will argue that Frankfurt has not offered a situation in which a factor eliminates the agent's ability to do otherwise without accounting for the performance of his action. I start, however, by presenting an objection to the premise 4 of his argument, which is, in my view, mistaken but an objection that will serve to clarify the nature of Frankfurt's argument.

2.2 The 'Locked Room'

An initial worry about Frankfurt's argument arises when we notice that there are factors which do not explain agents' behavior that are nevertheless relevant for their responsibility. A good illustration of this phenomenon is Locke's example (which has perhaps inspired Frankfurt's own example) of a man who voluntarily stayed in a room not knowing that there was no way for him to get out of it because the doors were locked. In this case, the fact that the doors were locked played no role in leading him to his decision to stay and to his staying in the room. But it seems that this fact is relevant to our judgment about his responsibility for staying in the room. For, it seems that the agent in this example is not responsible for not leaving the room or for staying in the room, but only for deciding to stay in the room, or not trying to leave it, exactly because the doors of the room were locked.

David Widerker and some other philosophers have argued on the basis of this phenomenon that the premise 3 of in Frankfurt's argument is false. That is, they argued that it is false that if something does not explain why an agent performed an action, it has no bearing on agent's responsibility for that action. However, I don't think that this phenomenon can serve this purpose.[37] In my view, this phenomenon is not a problem for Frankfurt's argument because his argument is focused, or rather should be focused, only on responsibility for *basic* mental actions such as deciding and willing and this phenomenon does not occur when these actions are concerned. This is so because the phenomenon in question is a consequence of the fact that the performance of non-basic actions depends on the cooperation of the agent's environment. The factors in our environment usually don't explain why we have decided to perform actions we performed but often influence *what* actions we actually

[37]I will argue in the final section of this chapter that it is not advisable for the proponents of PAP to use this phenomenon in defense against Frankfurt's argument. Here I want to explain why the phenomenon in question does not represent a problem for the Frankfurt's argument.

perform and what actions we are able to perform by influencing the results of our basic actions. Thereby they bear on *what* we are responsible for. In particular, they bear on our responsibility for actions by determining what actions we perform when we decide to perform an action, and bear on our responsibility for omissions by determining what actions we would be responsible for if we decided and tried to perform them. But clearly they *cannot* have this sort of influence on our responsibility for basic actions of deciding and trying because these actions are not performed by other 'decidings' or 'tryings'.

It would not be inaccurate to say that these factors influence our responsibility for non-basic actions by influencing our ability to perform those actions.[38] But thereby they are influencing only *what* we are responsible for and not *whether* we are responsible. Whether we are responsible or not depends on our decisions and tryings. But I cannot see how some factor that does not explain why we made some decision or tried to do something could have a bearing on the claim that we have done that responsibly (A factor which does not explain why we made a certain decision could explain, though, the content of our responsibility in making that decision, e.g. whether we are *morally* responsible or not[39]).

Therefore, it is plausible to consider Frankfurt-style examples as counterexamples to the claim that our responsibility for our *basic* acts, or simply that our responsibility, depends on our having ability to do otherwise. In order to show that Frankfurt's argument against PAP fails it is necessary to show that his example is incoherent. In other words, it is necessary to show that it is impossible that the agent is responsible although he could not have decided otherwise.

[38] This, however, does not confirm PAP. For, ability at stake here can be compatible with the agent's lack of ability to do otherwise because the agent lacks the ability to choose to do otherwise.

[39] An example of such a factor is awareness that doing something would be wrong. For such awareness may not play any role in one's acting wrongly, but it would certainly be relevant for the fact that one is morally responsible. See David Widerker, "Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 61-62.

2.3 The Compatibilists' Answer to Frankfurt's Argument

At first sight, compatibilist and incompatibilist who believe in PAP are "in the same boat" when it comes to Frankfurt-type examples. There is no doubt that Frankfurt thought so when he presented his example. For, as I mentioned, in his article he claims that his example can be modified so that Jones lacks ability to do otherwise on *any* account of ability. However, closer examination shows that it is not so. For, if the notion of ability is understood in the 'compatibilist way,' Frankfurt-type stories do not pose a threat to PAP. Moreover, if some Frankfurt-type story is coherent it provides an unexpected resource in showing that ability to do otherwise is compatible with determinism. How is that possible?

As I mentioned in the previous chapter, incompatibilists think that for an agent to be able to perform a particular action (in the actual circumstances) the action must be compatible with all the actual facts. For that reason, Frankfurt's example poses a challenge to incompatibilists because Black's presence is incompatible with Jones's doing otherwise.. But compatibilists do not face the same difficulty because they don't think that an agent can do something only if the action is compatible with *all* the actual facts. Some actual facts such as the facts about the laws of nature or the past are not relevant, in their view, for the question about the agent's ability to do otherwise. So, there is conceptual space for them to argue that the agents in the Frankfurt-type stories have the ability to do otherwise even if the exercise of that ability is impossible due to the presence of the intervener. Still, they need to explain how that can be so. That is, they have to explain why the presence of Black does not eliminate Jones's ability to do otherwise.

Interestingly, an explanation of this sort has emerged from considerations of some difficulties with the traditional compatibilist account of ability to do otherwise: the

conditional analysis of ability. According to that analysis, the meaning of statements about unexercised abilities is to be analyzed in terms of counterfactual conditionals. Thus, to say that someone has the ability to raise a hand means (among other things) that if the person decided (or wanted) to raise a hand, she would raise it. Understood in this way, ascription of unexercised abilities is compatible with determinism, for it can be true that the person would exercise the ability if she decided, even if it is determined that person *will* not decide to exercise it and will not exercise it.

One of the main objections to this analysis of ability to do otherwise and to conditional analysis of powers and dispositions in general is that we may truly ascribe a power or a disposition to an object even when the corresponding conditional is false and vice versa. Here is an example of this phenomenon. According to the conditional analysis a wire is live, that is, it is disposed to conduct electric current (although it does not conduct it) if it would conduct it were it attached to the source of the current. It is conceivable, however, that a live wire would not conduct electric current upon touching a source of electricity because of the presence of a device called 'fink' which would change its the inner structure upon touching to the source of electricity. This is called the phenomenon of 'finkish disposition.' This phenomenon seems to show that the truth of a conditional is *not necessary* for the possession of the corresponding ability. On the other hand, the phenomenon of 'finkish lack of disposition,' seems to show that the truth of the conditional is *not sufficient* for the truth of the corresponding ability claim. For, it is conceivable that a dead wire would conduct electric current upon touching the source of electricity, because the fink would make it alive instantaneously in those circumstances.

Many philosophers were convinced by this and other problems that conditional analysis of abilities should be abandoned. However, some philosophers concluded instead that ability claims cannot be analyzed in terms of *simple* conditionals. The first who

suggested a revised, more complicated conditional analysis of ability claims was David Lewis. Key to his suggestion and all subsequent ones was that the analysis in terms of simple conditionals should be supplemented with the requirement that the object does not change with respect to its power in the circumstances in which it should manifest it. More precisely, since Lewis identified dispositions with intrinsic features of objects in virtue of which they manifest dispositions in the relevant circumstances, he added the condition that the thing does not change with respect to those intrinsic properties.[40]

This solved the conditional analysis' problem with finkish properties. For, according to the new analysis, the wire is live before it was attached to a source of electric current because it would conduct electricity if it *retained* the ability to conduct it (that is, pace Lewis, if it retained its intrinsic properties). The revision also solved the problem with the finkish lack of dispositions. For, it is true that a dead wire would not conduct electric current if connected to a source of electricity and did not lose its disposition to conduct electric current, that is, according to Lewis, if it retained its intrinsic features.

But how is this relevant to Frankfurt-type examples? Kadri Vihvelin explains this in the following passage:

> Frankfurt's argument fails because Black is a fink - a superfink. Black's presence makes it the case that *all* of Jones' abilities, including the abilities which constitute free will, are finkish. Black leaves all of Jones' abilities intact, but Black's power and intentions ensure that if Jones ever begins or tries to exercise any of his abilities in any way contrary to Black's intentions, he will immediately lose that ability.[41]

In other words, the compatibilists' response to Frankfurt goes like this. Since Black is a sort of fink, just as we can say that a wire has the power to conduct electricity even though it

[40]Kadri Vihvelin follows Lewis in applying his account of dispositions to the problem of free will. Ferenc Huoranszki, however, offers a slightly different account which does not require identification of free will with intrinsic properties of the agent. See Ferenc Huoranszki, *Freedom of the Will: A Conditional Analysis* (New York: Routledge, 2011), 83-95.

[41]Kadri Vihvelin, "Free Will Demystified," *Philosophical Topics* 32 (2004):

would not do that in the circumstances in which it should conduct it, we can say that Jones has the ability to do otherwise even though he would never exercise it in Black's presence. For just like the fink attached to a wire, Black is inactive when Jones is not about to exercise his power to do otherwise, but eliminates his power when he is about to exercise it.

Therefore, Frankfurt's claim that his argument does not depend on the particular account of ability is false. For if the ability to do otherwise is understood in the way compatibilists have traditionally suggested, the argument completely loses its plausibility. Moreover, it provides support for compatibilism of ability to do otherwise because it shows that the claim that an agent can do otherwise even though it is impossible that he do otherwise is not an ad hoc hypothesis introduced only for saving compatibilism.

However, there are two problems with this reply. First is that the compatibilist understanding of ability or the conditional analysis may be an incorrect account of the meaning of ability to do otherwise. For even if it avoids the objection based on the possibility of finkish dispositions or lack of such dispositions, the conditional analysis proponent may not have a good answer to other objections. Second, more important problem in the present context concerns the claim that Black acts like a fink in the sense that he can prevent Jones from exercising his ability to do otherwise when he is about to exercise it without doing anything to the agent. In my view, this claim is false, for reason that I will present in the section 2.6. In the next section I consider an objection to his argument which is widely considered as unsuccessful but which is important for proper understanding of 'alternative possibilities,' relevant to moral responsibility.

2.4 The 'Flicker of Freedom' Strategy

As we have seen, the ability to do otherwise which is at stake in Frankfurt's argument is essentially related to the conceivability of agent's doing otherwise in the presence of the counterfactual intervener[42]. Thus, Jones can do otherwise if it is conceivable, given Black's presence, that he performs an alternative action or refrains from what he is actually doing. Frankfurt suggests, and his compatibilist opponents agree, that this may not be conceivable. However, a careful look at what goes on in the alternative sequence of his example (the sequence in which Back intervenes) shows that Jones has some alternative to what he is doing. In that sequence, it becomes clear to Black (who is an excellent judge of such things) that Jones will decide otherwise. But, Frankfurt does not say how that happens. One possibility is that Black notices Jones's beginning to make an alternative decision. But, in that case, Jones has the alternative possibility to *begin* to make an alternative decision.[43] Another possibility is that Black accurately predicts that Jones will decide otherwise on the basis of noticing a twitch, a blush, or a neurological pattern that Jones involuntarily emits every time he is about to begin to decide not to perform the action Black wants him to perform. In that case, Black can intervene before Jones even begins to make an alternative decision. However, Jones would still have an alternative: he would be able to emit the sign that would trigger Black's intervention.[44]

There are, however, other sorts of alternatives in Frankfurt's example which have nothing to do with Black's predictive powers. That is, it is possible to find an alternative in Frankfurt's example even if there is nothing on the basis of which Black makes a decision to intervene. One sort of alternative possibility with that characteristic 'becomes visible' when we consider Frankfurt's example in the light of the theory of identity of events (assuming that

[42] For it is assumed that without the presence of Black there is no reason to think that Jones cannot do otherwise.
[43] See John Martin Fischer, "Responsibility and Alternative Possibilities," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006): 31.
[44] As we shall see, examples of this sort play a very important role in the dialectic of the discussion and have come to be known called 'prior sign Frankfurt examples.'

actions are events) which says that causes of events are essential to their identity. On this theory of event identification, Jones would perform the same type action in the alternative scenario as he actually performs, but he would not perform the same particular action due to the presence of Black's intervention in the causal origin of that action.[45] Thus, on this view, in the actual scenario Jones has at least this alternative possibility to perform a different particular action.

A somewhat similar alternative possibility appears when we think about the exact content of Jones's responsibility. According to Frankfurt, Jones is responsible for doing what he does, which is what Black wants him to do. However, it may be more accurate to say that Jones is responsible for doing that *on his own*. But if that is what Jones is really responsible for he has the following alternative possibility: he can do what Black wants him to do because of Black's intervention rather than on his own.

Finally, an alternative possibility that is available to Jones emerges when we consider Frankfurt's example through the lenses of the agent-causal libertarian theory. According to that theory, an event is an action if and only if it caused by the agent or if there is a chain of mental and physical causes that can be traced back to the agent. In addition, if the agent is to be truly free and responsible, the causal chain must not extend further than the agent. In other words, to have free will and be morally responsible a free agent must be a first cause uncaused. Now, if we assume that Jones is such a cause, Jones has the alternative possibility not to be the agent-cause the action he is actually performing (or is about to perform).

Therefore, if Frankfurt's argument against PAP requires a case in which an agent lacks any alternative possibility whatsoever, the argument fails. For, it seems impossible to

[45] The distinction between the type of action and particular instantiation of that type of action is important also because the agent may not be able to avoid performing a certain type of action because the intervener would actually intervene at some moment. In that case, the agent may not be able to refrain from performing the action of a given type, but he may be able to refrain from performing a particular instantiation of it, that is, he may refrain from performing it at a certain time.

modify Frankfurt's example so that Jones lacks *any* alternative possibility whatsoever, given that the presence of at least some of the above mentioned alternatives does not depend on the special details of a particular Frankfurt-style example, that is, it does not depend on the kind of mechanism that is supposed to eliminate alternative possibilities.

However, does Frankfurt's argument really require the lack of alternatives of the sorts just mentioned? It does not seem so because the alternatives of these sorts don't seem to be able to ground moral responsibility. For, as Fischer points out, these alternative possibilities or flickers as Fischer calls them, don't seem *robust* enough to ground moral responsibility. For it is difficult to see how having them could contribute to the agent's control over his actions. Each of the four types of flickers I have mentioned is manifested in some involuntary and unconscious behavior. In the first case, the agent can involuntarily emit a certain sign. In the second, he can perform a different particular action, but he cannot do that voluntarily. Finally, he can do something but not on his own. However, the power not to do something on one's own is obviously not a power to do something voluntarily. The same is true of the power not to do something if not doing it can only be the result of external factors unknown to the agent.

But, even so these flickers may be exactly what we need in order to be morally responsible for what we do. Metaphorically speaking, perhaps they provide all the elbow room we need in order to control our actions. And it may be that determinism poses a threat to moral responsibility exactly because it eliminates these flickers. Some (or perhaps many) philosophers think that this is indeed the case, although not because these alternatives per se ground moral responsibility but because they are necessary byproducts[46] of something else that grounds moral responsibility. Thus, according to incompatibilists who think that what really matters for moral responsibility is what goes on in the actual sequence leading to

[46] I am not sure, though, that 'byproduct' is the right word here. Maybe it would be more accurate to say 'necessary condition.'

action, these flickers are relevant because without them the actual sequence could not be indeterministic and without that the agent could not be morally responsible.

However, even if this is true, the flicker of freedom strategy, (the strategy of arguing against cannot save PAP from Frankfurt's argument. For, according to PAP, alternative possibilities are relevant for moral responsibility *per se* and not because they indicate that some other condition is satisfied. But the failure of this strategy is very instructive. It helps us to understand better the notion of alternative possibility which is at stake in the debate about Frankfurt's argument. The alternatives that matter in this debate are actions the agents can perform voluntarily, or perhaps more accurately, which don't just happen to them, which are under their control. Thus, proponents of Frankfurt's argument must present a case in which an agent lacks all such alternatives but because of some factor which does not actually explain why he performs that action.[47]

In the following section I will present a strategy for defending PAP, which in my view, shows that there must be such alternatives in every version of Frankfurt's example in which the agent acts responsibly. Then, I will present some modified versions of Frankfurt's example and argue that in spite of their authors' ingenuity they fail to show that the strategy in question is unsuccessful.

2.5 The 'Dilemma Defense'

At the beginning of the previous section I pointed to the lack of explanation of Black's ability to predict what Jones will decide in the Frankfurt's example. I then noticed that this fact leaves room for the suggestion that Jones may begin to make an alternative decision. I added, however, that the alternative possibility of that sort can be eliminated by

[47] This is often called 'IRR situation'.

introducing a prior (involuntary) sign in the example which tells Black that Jones will decide otherwise if left on his own (and which serves as a triggering event). As we have seen, the prior sign lacks robustness necessary for an alternative possibility to be relevant for moral responsibility because it is involuntary. But, the fact that Black needs a prior sign to intervene is important for a completely different, and in my view, much more powerful objection to Frankfurt's argument presented first by Robert Kane and developed by David Widerker. The objection has a form of the dilemma: Jones's decision is either preceded by a sign (or the absence of a sign) which *guarantees* that Jones will do what Black want him to do, or it is not preceded by such a sign. If former is the case, it is not uncontroversial that Jones is morally responsible. For, this conclusion will be unacceptable for incompatibilists. But, if latter is the case, that is, if Jones is a libertarian free agent, Black cannot know when and what will Jones decide to do if left on his own. So, in that case he has two options: he can either wait to see what Jones will decide and if he begins to decide not to do what he wants him to do force him to decide and do otherwise; or he can intervene without waiting for the sign. But, if he chooses the former option his intervention will come too late, for beginning to decide otherwise seems like a robust alternative possibility (it is something that can be done voluntarily), whereas if he chooses the latter option it would no longer be true that Jones has made the decision he wants him to make on his own.

It is widely considered that this is the most powerful objection to Frankfurt's argument against PAP. However, many philosophers have tried to refute it by presenting new more elaborate Frankfurt-type examples especially tailored for incompatibilists, that is, examples which do not presuppose determinism. In what follows I will consider four versions of such Frankfurt-type examples and argue that in spite of their initial appeal none of them shows an agent who acts responsibly without having the ability to decide to do otherwise.

2.5.1 Stump's Example

In response to the dilemma objection Eleonore Stump has devised a more elaborate version Frankfurt's example, offering more details in particular about how "the fictional coercive mechanism works and what it operates on."[48] In fact, her example is a revised version of Frankfurt-type example presented earlier by John Martin Fischer. She named it G after a neurosurgeon that plays the role of the counterfactual intervener in her example. The example goes like this:

> (G) Suppose that a neurosurgeon Grey wants his patient Jones to vote for Republicans in the upcoming election. Grey has a neuroscope which lets him both observe and bring about neural firings which correlate with acts of will on Jones's part. Through his neuroscope, Grey ascertains that every time Jones wills to vote for Republican candidates, that act of his will correlates with the completion of a sequence of neural firings in Jones's brain that always includes, near its beginning, the firing of neurons a, b, c (call this neural sequence 'R'). On the other hand, Jones's willing to vote for Democratic candidates is correlated with the completion of a different neural sequence that always includes, near its beginning, the firing of neurons x, y, z, none of which is the same as those in neural sequence R (call this neural sequence 'D'). For simplicity's sake, suppose that neither neural sequence R nor neural sequence D is also correlated with any further set of mental acts. Again for simplicity's sake, suppose that Jones's only relevant options are an act of will to vote for Republicans or an act of will to vote for Democrats.
> Then Grey can tune his neuroscope accordingly. Whenever the neuroscope detects the firing x, y, and z, the initial neurons of neural sequence D, the neuroscope immediately disrupts the neural sequence, so that it isn't brought to completion. The neuroscope activates then the coercive neurological mechanism which fires the neurons of neural sequence R, thereby bringing it about that Jones wills to vote for Republicans. But if the neuroscope detects the firing of a, b, and c, the initial neurons in neural sequence R, which is correlated with the act of will to vote for Republicans, then the neuroscope does not interrupt that neural sequence. It doesn't activate the coercive neurological mechanism, and neural sequence R continues, culminating in Jones's willing to vote for Republicans, without Jones's being caused to will in this way by Grey.

[48] Eleonore Stump, "Moral Responsibility without Alternative Possibilities," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 140.

> And suppose that in (G) Grey does not act to bring about neural sequence R, but that Jones wills to vote for Republicans without Grey coercing him to do so.[49]

This example does not include a sign which *precedes* the agent's action and guarantees that that agent will perform the action. So, prima facie, there is no reason to think that the example begs a question against the incompatibilist.[50] The example features a sign that tells the counterfactual intervener Grey (or his neuroscope) when to intervene, but the sign occurs simultaneously with the action he wants to prevent. For, the sign is part of the physical (neural) basis or correlate of the (alternative) mental act, i.e. the decision to vote for Democrats. However, according to Stump, even though the decision may be indeterministic, Grey's intervention would not have to come too late, (when Jones has already made a decision or started to make the alternative decision), as the dilemma objection predicts. For, according to Stump, the alternative decision corresponds to the completed sequence of neural firings and does not occur at all if the sequence is interrupted at any point. Nevertheless, since Grey actually does not interfere into Jones's deliberation process, intuitively Jones is morally responsible for his decision.

If this story is coherent Stump has found a way around the dilemma objection. That is, she has presented a case in which a Frankfurt-intervener eliminates all *robust* alternative possibilities without rendering the agent's choice causally determined. But is her story really coherent? The answer to this question, in my view, depends on what we should say about her account of the relation between free decisions and their neural correlates. In particular, it depends on her claim that (free) decisions are correlated with sequences of neural events

[49] Stump, "Moral Responsibility without Alternative Possibilities," 140.

[50] This is not true, though, if alternative decision corresponds only to the completion of the sequence of neural firings, that is, o the firing of the last neuron in the sequence. This has been observed by David Widerker. However, Stump has made it clear in a response to Widerker that the decision can also correspond to the entire sequence. In that case, the problem that the example will be unacceptable to incompatibilists does not occur. I have assumed this interpretation of Stump's example in the text.

extended in time and occur only when those sequences are completed. I find these claims problematic because it is difficult to see how we could control our behavior if they were true. For, if our decisions correspond to sequences of neural firings and occur only when those sequences are completed it seems that our decisions are 'brewing' in our brains before we are aware of them. But, in that case, our decisions would seem to be mere epiphenomena. Second, it is not clear that mental acts must be correlated with neural events that are extended in time. Why couldn't the relevant neural events be instantaneous just like decisions seem to be introspectively? That they cannot be instantaneous follows from Stump's assumption that mental acts correspond to *sequences* of neural firings. But, why couldn't they correspond simply to *simultaneous firings* of a certain numbers of neurons? I don't see how Stump could reply to these important questions. Therefore, I think she has not offered a good defense of Frankfurt's argument from the dilemma objection.

2.5.2 Hunt's Example

As we have seen, according to the dilemma objection Frankfurt-style examples must feature a (prior) sign because without such a sign the counterfactual intervener would not know when to intervene; and without that knowledge he can ensure that the agent will do what he wants him to do only by actually forcing him to do that. That is, without a sign the intervener could not be a merely counterfactual intervener and he would have to interfere with the agent's actual deliberation. However, perhaps the intervener does not have to be a merely counterfactual intervener (who actively eliminates alternatives only in the counterfactual sequence, i.e. when the agent is about to do otherwise) in order to eliminate alternative possibilities without actually interfering with the agent's deliberative process. For, it seems possible that the intervener actually eliminates all alternatives possibilities without

thereby influencing the agent's decision (it could do that, for instance by sheer coincidence). David P. Hunt illustrates this possibility with the following example:

> Suppose the driving instructor can lock his wheel at a certain position to prevent the student driver from steering beyond that range, and Black has placed a 'left lock' on his steering wheel to block the possibility that Jones might take the road to the left; Jones, however, bears right at the fork and never encounters the lock. The principal difference between this kind of case and the one involving the counterfactual alternative-eliminator is that the passive eliminator is in place in the actual world, though the sequence of events actually productive of Smith's death never intersects with it (hence its 'passiveness'). But the moral it conveys appears to be the same. A steering lock is no less effective than is Black counterfactual resolve in ensuring that the car is going to hit Smith and that there is nothing Jones can do to avoid this outcome. Moreover, there is no less reason in this case to regard Jones as a free agent in killing Smith. The passive alternative-eliminator does not figure in the actual sequence; in its absence, Jones would have done everything the same. If these reasons support Jones's free agency in the face of a counterfactual alternative-eliminator, they equally support his free agency when a passive alternative-eliminator is at work.[51]

In this example Jones obviously has alternative possibilities: he can decide and try to steer the car in the alternative way. However, in Hunt's view, "there is no reason to think that these alternatives cannot be eliminated in the same way (and with the same consequences for Jones's free agency): the relationship between Jones and the car's direction appears to model the relationship between an agent and any action of that agent, no matter how immediate."[52] In other words, Hunt claims that his example can be a model for a successful Frankfurt-style case. For obvious reasons, the method of eliminating alternative possibilities in this model is usually referred to as a 'blockage.'[53]

According to Hunt, an advantage of this method of eliminating alternative possibilities in which the alternative possibilities are eliminated by what he calls a 'passive

[51] David P. Hunt, "Freedom, Foreknowledge and Frankfurt," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 170.
[52] Hunt, "Freedom, Foreknowledge and Frankfurt," 171.
[53] The name, I think, comes from Fischer.

alternative-eliminator' over that involving 'counterfactual alternative-eliminator' is that the former eliminator is much more effective in expunging the alternatives than the latter. For, as Hunt explains,

> …there is an upper limit on how far the counterfactual alternative-eliminator can be tightened, since its triggering structure requires that *some* alternative be accessed before the mechanism comes 'on line.' There is, on the other hand, no evident upper limit on the restrictions imposed by a passive alternative-eliminator. Any alternative can be passively blocked; and because the alternative is eliminated passively, the actual sequence, along with Jones's free agency, is unaffected.[54]

However, as Robert Kane points out, there is also an upper limit on how much the passive alternative-eliminator "can be tightened." For it seems impossible to eliminate all alternative possibilities without thereby causally determining the action in question. Robert Kane explains this in the following passage:

> In [a case in which every other alternative is blocked except the agent's choosing *A* at *t*], of course, there *are* no alternative possibilities left to the agent; every one is blocked except the agent's choosing *A* at *t*. But now we seem to have determinism pure and simple. By implanting the mechanism in this fashion, a controller would have predetermined exactly what the agent would do (and when); and, as a consequence, the controller, not the agent, would be ultimately responsible for the outcome. Blockage by a controller that rules out all relevant alternative possibilities is simply predestination; and on my view at least predestination runs afoul of ultimate responsibility.[55]

Hunt's reply to this objection is that although the passive alternative-eliminator may causally determine Jones's action, it does not mean that Jones murders Smith "because of the alternative-eliminator,"[56] that is, because of that causal determination. Jones could still

[54] Ibid.
[55] Robert Kane, "Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson," *Philosophy and Phenomenological Research* 60 (2000): 162.
[56] Hunt, "Freedom, Foreknowledge and Frankfurt," 173.

murder Smith because of his own "deliberations, decisions, intentions and so on."[57] Hunt explains the relevance of this point in the following way:

> The key Frankfurtian insight is that what happens in the actual sequence is all that's relevant t judgments of free agency and moral responsibility. If Jones's murder of Smith, along with such crucial preliminaries as Jones's decision to murder Smith, are determined by a causal chain operating within the actual sequence itself, then the libertarian must deny that Jones is functioning as a free agent; but if Jones's agentially relevant states are determined by causal factors operating outside the actual sequence, the libertarian who has taken Frankfurt's critique to heart might well deny that such causal determinism counts against Jones's freedom.[58]

However, it is not clear how it is possible that "Jones's agentially relevant states are determined by causal factors operating outside the actual sequence." If a factor causally determines an action must be part of the actual sequence. It is possible, of course, that the agent does not perform the action only because of that factor. However, libertarians would generally not willing to say that the agent is free and responsible just because of that.

I think, thus, that Hunt's blockage strategy also fails to meet the dilemma objection. However, Alfred Mele and David Robb have developed a more sophisticated Frankfurt-style example which uses (together with another strategy) a version of blockage which seemingly avoids the worry about begging the question against incompatibilists.

2.5.3 Mele and Robb's example

Mele and Robb present their Frankfurt-style example in the following way:

> Our scenario features an agent, Bob, who inhabits a world at which determinism is false … At *t1*, Black initiates a certain deterministic process *P*

[57] Ibid. 173.
[58] Ibid. 172-173.

> in Bob's brain with the intention of thereby causing Bob to decide at *t2* (an hour later, say) to steal Ann's car. The process, which is screened of from Bob's consciousness, will deterministically culminate in Bob'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, for example he is dead by *t2*) … The process is in no way sensitive to any 'sign' of what Bob will decide. As it happens, at *t2* Bob decides on his own to steal the car, on the basis of his own indeterministic deliberation 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 Bob's decision.[59]

Mele and Robb clarify that they identify the neural events that are correlated with decisions with 'lighting up' of 'decision nodes.'

> The 'lighting up' of node *N1*represents his deciding to steal the car, and the 'lighting up' of node *N2* represents his deciding *not* to steal the car. Under normal circumstances and in the absences of preemption, a process's 'hitting' a decision node in Bob 'lights up' that node. If it were to be the case both that *P* hits *N1*at t2 and that *x* does not hit *N1* at t2, then *P* would light up *N1*. If both processes were to hit *N1* at *t2*, Bob's indeterministic deliberative process, *x*, would light up *N1* and *P* would not.[60]

Finally, they explain what happens when the process x and the process P 'diverge' so that the latter hits the node N1 and the former node N2.

> …if *x* and *P* were to 'diverge' at *t2*, so that *x* hits *N2* and *P* hits *N1*, *P* would light up *N1 and* x would not light up *N2*. Why? Because 'by *t2P* has neutralized all of the nodes in Bob for decisions that are contrary to a decision at *t2* to steal Ann's car …In convenient shorthand , by *t2P*has neutralized *N2* and all its "cognate decision nodes".[61]

What should we say about this case? Is it vulnerable to the objections raised against the other Frankfurt-style examples considered so far?

[59] Alfred Mele and David Robb, "Bbs, Magnets and Seesaws: The Metaphysics of Frankfurt-style Cases," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 128.
[60] Ibid. 129.
[61] Ibid.

At first sight, it may seem so because this example also involves blockage like the previous one; and blockage is problematic because it renders the action in question causally determined and because it involves interference with the process which actually leads to action. However, as Kane points out, there is a significant difference between this case and the case envisaged by Hunt. In Hunt's case *all* alternative possibilities are blocked. In this case the agent has some alternative possibilities, but none of them are robust, e.g. the agent might be incapable to decide because he is dead. For this reason, Kane calls the former type of blockage "pure blockage," and the one in Mele/Robb scenario the "modified blockage." As Kane points out, the main advantage of the modified blockage over the pure blockage is that the former does not render the actual sequence causally deterministic and thus does not expose Mele and Robb to the objection that their example begs the question against the incompatibilist.

Nevertheless, according to Kane, in the modified blockage just as in the pure blockage, Frankfurt's controllers actually interfere with the agent's process of deliberation. Kane explains this in the following passage:

> First, allowing some non-robust AP's does not change the situation regarding the crucial premise of Mele/Robb example. This crucial premise is that the controller's deterministic process *P* is 'causally isolated' from Bob's decision-making process *x* and 'in no way' interferes with Bob's decision-making process. This crucial premise remains false in the modified blockage case even when non-robust AP's are allowed. For in the modified scenario, the controller's process P must still block all *robust, voluntary* alternative possibilities of Bob's – *B, C, D* – at *t2*, if it is to do its job effectively; and this necessary blocking of robust alternatives still involves *actual* intervention by *P* in Bob's decision-making process at *t2*, even if *P* should leave some non-robust AP's at *t2*. The controller is no mere counterfactual intervener who does not actually intervene in the situation even in modified blockage cases. His actual intervention limits all of Bob's robust alternatives to one.[62]

[62] Robert Kane, "Responsibility, Indeterminism and Frankfurt-style Cases: A Reply to Mele and Robb," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 101.

I think that Kane is right that the controller must actually interfere with Bob's decision-making process in order to eliminate all robust alternative possibilities that Bob by assumption has in his absence. However, I also think that more needs to be said to make this reply convincing. In particular, it must be explained why blockage cannot be conceived so that there really is no contact between the actual neural process correlated with Bob's deliberation and the alternative-eliminating process P. This would be possible, for instance, if Stump were right that mental acts correspond to sequences of neural firings and that a mental act occurs only when the corresponding neural sequence is completed. The reply Mele and Robb give to this objection in the following passage suggests that they have a very similar view of the relation between the mind and the brain:

> Setting aside recurrent neural networks[63] and other neuroscientific considerations, a critic may claim that *P*'s neutralizing *N2* and all its cognate decision nodes without interfering in Bob's indeterministic deliberation process, *x*, is a *conceptual* impossibility. Our diagnosis is that such a critic misunderstands our case. Imagine a pinball machine that is subject to indeterministic forces. Dave has covered four of the five circular 100-point bumpers with plastic so that there is only one 100-point bumper the pinball can actually touch. Dave 'neutralized' these four bumpers, one might say. Plainly, in a possible scenario, the plastic coverings have no effects at all on how the pinball moves. For example, Al, using the machine's plunger, might shoot the pinball into the playing field, and it might bounce off the uncovered 100-point bumper and out of the exit hole without touching or being affected in any way by the plastic covers. Similarly, as far as we can see, there is no conceptual problem with the supposition that *P*'s neutralizing of *N2* and all its cognate nodes has no effect on what goes on in Bob's indeterministic process of deliberation.[64]

The key feature in this scenario seems to be that there is a space in which the pinball indeterministicaly moves. More precisely, although the pinball moves toward the uncovered 100-point bumper, it might swerve and move in the direction of some of the covered bumpers. This possibility of a swerve is analogous to the firing of neurons x, y, and z in

[63] The idea that brain processes correlated with mental process involve the activity of recurrent neural networks has been suggested by Robert Kane.
[64] Alfred Mele and David Robb, p. 132

Stump's example. But, as I mentioned earlier (and the pinball example perhaps makes it even clearer), if this is how the mind is related to the brain, it is very difficult to see how there could be free decisions. And, although the pinball example is just an illustration of an idea, if that idea were correct, I think our freedom of will would be no greater than the freedom of a pinball machine.

If this observation is correct there are serious conceptual problems for the blockage strategy. In other words, the dilemma objection thus seems to be an insurmountable obstacle for the proponent of Frankfurt's argument. But before drawing this conclusion one more example needs to be considered.

2.5.4 Pereboom's example

In the previous two sections I presented attempts to produce a successful Frankfurt-style example with no signs which help the counterfactual intervener or a mechanism he has set up to eliminate alternative possibilities. In this section I will present yet another 'prior-sign Frankfurt-style example.' However, this example does not contain a prior sign of the problematic sort. The example is due to Derk Pereboom. He calls it Tax Evasion. It goes like this:

> *Tax Evasion* (2): Joe is considering whether to claim a tax deduction for the substantial local registration fee 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 regardless of the cost to others, and no matter whether advancing his self-interest involves illegal activity. Crucially, his psychology is such that the only way in this situation he could fail to choose to evade taxes is for moral reasons. (The phrase *failing to choose to evade taxes* is meant to encompass not choosing to evade taxes and choosing not to evade taxes.) His psychology is not, for example, such that he could fail to choose to evade taxes for no reason or simply on a whim. In addition, it is causally necessary for his failing

> to choose to evade taxes in this situation that he attain a certain level of attentiveness to these moral reasons. He can secure this level of attentiveness voluntarily. However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, Joe could, with his libertarian free will, either refrain from choosing to evade taxes or refrain from so choosing (without the intervener's device in place). More generally, Joe is a libertarian free agent. But to ensure that he choose to evade taxes, a neuroscientist now implants a device, which, were it to sense the requisite level of attentiveness, would electronically stimulate his brain so that he would choose to evade taxes. In actual fact, he does not attain this level of attentiveness, and he chooses to evade taxes while the device remains idle.[65]

The prior sign in this case is the "attaining of certain level of attentiveness to moral reasons." Like other prior signs in Frankfurt-type cases, this sign tells the counterfactual intervener when to intervene, or simply triggers the reaction of a device that he has implanted in the agent's brain. However, this sign is *not sufficient* for Joe's performing or refraining from performing some action, but only *necessary* for his performing an *alternative* action. Thus, the presence (or absence) of this sign does not indicate that the sequence actually leading to Joe's decision is deterministic. Consequently, incompatibilsits have no reason to worry that Joe is not morally responsible because of the presence of such sign. Nevertheless, according to Pereboom, in these circumstances the counterfactual intervener can be no less effective in eliminating robust alternative possibilities as in the case where the sign is a sufficient condition of the action the agent actually performs. Therefore, this Frankfurt-style example seems immune to the Kane/Widerker objection.

Before evaluating this claim, I must give some clarifications about this example. As Pereboom points out, it is uncontroversial that Joe is morally responsible for deciding to evade taxes even by libertarian standards because at any moment before making that decision he can voluntarily attain the required level of attentiveness to moral reasons. But, according

[65] Derk Pereboom, "Source Incompatibilism and Alternative Possibilities," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 193.

to Pereboom, this is not a robust alternative possibility because robustness has an epistemic dimension, that is, there is an epistemic requirement for robustness of an alternative possibility which this alternative does not satisfy. The epistemic requirement in question is that the agent knows that by exercising it he or she would thereby avoid responsibility for what they actually do. But, according to Pereboom, this is not the case with Joe's possibility to voluntarily attain a certain level of attentiveness to moral reasons. In this respect Joe is the same as the agent who can avoid deciding to kill another person by voluntarily drink a cup of coffee which is unbeknownst to the agent poisoned and would instantly kill him. For, just like this person, Joe does not know that he would avoid responsibility for a decision to evade taxes by exercising that alternative possibility (he does not know about the intervener and believes that considering moral reasons is not sufficient for not deciding to evade taxes).

In my view, however, Joe does have a robust alternative possibility in this case. For, as Carl Ginet has noticed, it is crucial that Joe has the power to voluntarily consider moral reasons at *any* moment before making a decision on whether to evade taxes. In addition, Joe knows at the time of making a decision to evade taxes that by considering moral reasons he would avoid making a decision to evade taxes *at that moment*. In other words, by Pereboom's standards, at every moment before making a decision Joe has a robust alternative possibility not to decide to evade taxes and consider moral reasons instead at the next moment. So, the reason why Joe seems to be morally responsible for deciding to evade taxes is that he is responsible for deciding to evade taxes at a particular time.

Obviously, the distinction between responsibility for doing something simpliciter (or by a certain time) and doing something at a particular time is crucial to Ginet's objection. For, Joe has a robust alternative possibility only with respect to deciding to evade taxes at a particular time but not with respect to deciding to evade taxes simpliciter or doing that *by* a certain time (the time when he is forced to do that). But, there seem to be several problems

with this strategy. First, it seems that responsibility for doing something at a particular time entails responsibility simpliciter because doing something at a particular time entails doing that simpliciter. Second, since we have no reason to doubt that Joe would be responsible for deciding to evade taxes (simpliciter) in the absence of Black, and Black plays no role in Joe's decision to evade taxes, we seem to have no reason to doubt that Joe is responsible for his decision to evade taxes in Black's presence. Third, saying that Joe is responsible only for deciding to evade taxes at a particular time sounds like saying that Joe is responsible only for the timing of his decision. But Joe is certainly responsible for more than just that. Finally, it is perhaps possible to modify Tax Evasion so that Joe cannot even decide to evade taxes at a particular time but he still seems morally responsible for his decision.

The first problem is easy to deal with. For, it is not true that if one is responsible for one fact one must be responsible also for every fact which that fact entails. Thus, as van Inwagen points out, the fact that a person is responsible for killing someone does not entail that he is responsible for the fact that that someone was 'mortal.' Or as, Ginet notices according to David Palmer, "while a person may be morally responsible for being in a particular room, he is clearly not morally responsible for the entailed fact that he is within a one-million-mile radius of the center or the earth."[66]

The second problem is a bit more difficult to eliminate. One could eliminate it by saying that even in the absence of Black Joe would only be responsible for deciding to evade taxes at a particular time because we are in general only responsible for doing things at certain times rather than just for doing them (or for doing them by a certain time). One could support this claim by pointing out that we can only be responsible for particular actions and particular actions occur at particular times. This would also eliminate the third problem because it would clarify that Joe is not responsible only for the timing of his action.

[66] David Palmer, "The Timing Objection to the Frankfurt Cases," *Erkenntnis* 78 (2013): 1017.

However, the problem with this reply is that it is more accurate to say that we are responsible for particular events under certain descriptions. And the description under which we are responsible for a particular event may not include the time of its occurrence. For instance, a person who caused another person's death by planting a landmine is responsible for killing that person, or simply for killing someone, but not for killing someone at a particular time, unless she knew or could have known that someone will step on the mine at that time.

This problem can be eliminated, in my view, by restricting the principle that we are responsible only for doing things at particular times to basic action such as decisions or tryings. This restriction is plausible because we are always in control of the time of making our decisions. It can never happen that a person is not aware of the timing of his or her decision as it can happen that a person is not aware of the timing of his or her non-basic action, for instance, as it is the case with a person who killed another person by planting a landmine (if the time of killing is determined by the other person was killed). Furthermore, this restriction makes sense because, as I have argued at the beginning of this chapter, Frankfurt-style examples concern only basic actions.

What I said here no doubt sounds a bit complicated and perhaps confusing. But if someone is not convinced by what I have said, I suggest that we notice that Frankfurt-style examples plausibly concern only the question about responsibility and not about the content of responsibility. In addition, we should notice that denying that the ability to perform action at a particular time would render discussion over Frankfurt-style examples meaningless. For the time of interveners intervention would be irrelevant. In any case it would be true that the agent could avoid the action the intervener has made him perform at some later time. And it would have very implausible consequences judgment about one's responsibility could vary depending on what happens after the action.

Thus, to succeed, Pereboom's example would have to show that Joe could be responsible for deciding to evade taxes at a particular time even though he could not avoid doing that *at that time*. Pereboom offers a modified version of Tax Evasion which supposedly shows that. He calls it the Tax Cut. In this example Joe is in a voting booth and is deciding between voting for a tax cut or against the tax cut. Joe has to decide either for or against the tax cut by pressing either yes or no button within a two minutes interval, otherwise he would have to pay a fine. The fine is "substantial enough so that in his situation he is committed with certainty to voting (either for or against), and this is underlain by the fact that the prospect of the fine, together with background conditions, causally determines him to vote." [67]A necessary but not sufficient condition for Joe's pressing the no button is that he imagines vividly that his boss would found out about his political views and decide not to promote him and he can do that either voluntarily or it can happen to him involuntarily. This is not sufficient condition for his pressing the no button because even if Joe imagines vividly that his boss finds out about his political views he could decide to press either yes or no button, that is, at every moment Joe can use his libertarian free will. That is, Joe could press either button "without the intervener's device in place."[68] However, the neuroscientist Black is again behind the scene, just this time his goal is to ensure that Joe will choose to press and press the yes button. Black has unbeknownst to Joe

> implanted a device in his brain, which, were it to sense his vividly imagining the no-promotion scenario, would stimulate his brain so as to causally determine the decision to vote for the tax cut. Joe does not exercise his imagination in this way, and he decides to vote in favor while the device remains idle.[69]

[67] Derk Pereboom, "Optimistic Skepticism about Free Will," in *The Philosophy of Free Will: Selected Contemporary Readings*, ed. Paul Russell and Oisin Deer, New York: Oxford University Press (2012): 15.
[68]Ibid.
[69]Ibid. 16.

Now, if we assume that Joe did not imagine vividly his boss finding out about how he voted and the picture of his boss finding out not to occurring involuntarily to him during the two minutes interval and imagine Joe deciding at the last moment within this interval, call it t2, to vote for the tax cut. According to Pereboom, Joe would be morally responsible for choosing for the tax cut at t2, even though he would not have a robust alternative possibility to doing that at t2. Joe would not have a robust alternative possibility because his commitment to deciding within the 2 minutes interval and his awareness that by imagining his boss finding about his voting decision at t2 he would fail to make a decision guarantee his decision to vote for the tax cut at t2, that is, his pressing the yes button at that time. Nevertheless, according to Pereboom, Joe would be morally responsible for pressing the button at t2 (pressing the button also had something to do with moral reasons).

However, I disagree with Pereboom. For, I don't think that what he says is acceptable from the libertarian perspective. This is so because in his example Joe is either causally determined to vote for tax cut at t2 or he is not causally determined to.[70] If he is causally determined he is not morally responsible for that according to libertarian standards. But if he is not, the libertarian would have no reason to think that Joe cannot do otherwise. Perhaps Pereboom had in mind psychological determination. Perhaps what he had in mind is that Joe could not do otherwise because of his own commitment. Nevertheless, I doubt that any libertarian would agree that Joe is directly responsible in that case even though he cannot do otherwise. Thus, I think that Pereboom's attempt to show that PAP is false fails.

2.6 Conclusion

[70] Palmer notices this in a footnote. He puts more emphasis on the claim that Joe's responsibility in this case is only derivative. I think, however, that this reply is the best reply and Palmer should have put more emphasis on it. Palmer, "The Timing Objection to the Frankfurt Cases," 1020.

In this chapter I defended the traditional view that free will requires ability to do otherwise from the powerful attack presented by Harry Frankfurt based on the idea that something may deprive the agent of all robust alternative possibilities without explaining why his actual behavior. I showed, I believe, that both traditional compatibilists and traditional incompatibilists have a plausible answer to Frankfurt's attack. Compatibilists can defend their version of the traditional view by pointing out that lack of alternatives in Frankfurt-style cases does not entail the lack of ability to do otherwise. Incompatibilists, on the other hand, can defend their view by showing that we cannot conceive of an example in which something deprives the agent of all robust alternative possibilities without explaining his actual behavior. My main goal, however, was to defend the incompatibilist answer to Frankfurt. In particular, I argued that the most prominent versions of Frankfurt's argument based on new Frankfurt-type scenarios are powerless against an objection raised by incompatibilists Robert Kane and David Widerker. I argued that those versions of the argument fail either because they presuppose an implausible theory of mind (Stump's argument and Mele and Robb's argument), or because they beg the question against the incompatibilist (Hunt's argument), or because they presuppose a mistaken conception of responsibility and of the role of Frankfurt-type examples (Pereboom's argument).

Thus, although I cannot exclude the possibility that someone will come up in the future with a successful Frankfurt-style attack on the traditional view that moral responsibility requires alternative possibilities, I think that all of the existing Frankfurt-style attacks on this view fail. Nevertheless, I think that Frankfurt's argument deserves place it has in the free will debate because it has encouraged philosophers to question the idea that free will requires ability to do otherwise. In chapter 4, I will consider another, in my view, much more promising attempt to show that the idea is false.

CHAPTER 3: LIBERTARIAN THEORIES OF FREE WILL

In the first chapter I argued that determinism poses a threat to the ability to do otherwise. If this is so and if free will requires ability to do otherwise, determinism undermines free will. So, it seems that free will can exist only if determinism is false. Libertarians about free will think that this is indeed the case. They think that free will is incompatible with determinism but not with indeterminism. In addition, they think that indeterminism is true in the actual world and that some human beings actually are free and responsible agents. In other words, libertarians not only believe that free will is *possible*, but believe that free will actually *exists*.

However, according to an old tradition in the free will debate, indeterminism is just as inhospitable, if not even more inhospitable, to free will and responsibility than determinism. The challenge which indeterminism represents to libertarianism becomes visible when we focus on the issue of control. It is clear that without a sufficient degree of control we cannot be morally responsible for what we do. Control is in turn the reason why moral responsibility requires ability to do otherwise. For, it seems that we have control over what we do only if we can do otherwise.[71] But, while indeterminism secures ability to do otherwise or openness of alternatives, which seem necessary for control, it seems to be incompatible with control. For, undetermined events seem to be the result of chance and chance events cannot be under

[71]John Martin Fischer, however, argues that a variety of control which is required for moral responsibility – 'guidance control' - does not involve ability to do otherwise. In his view, ability to do otherwise is important for 'regulative control', but that kind of control is irrelevant for moral responsibility. See John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, (Cambridge: Cambridge University Press, 1998)

anyone's control.[72] If this is so, indeterminism is incompatible with moral responsibility because it undermines control.

The worry that indeterminism undermines control was for a long time a very strong motive for accepting compatibilism. For instance, David Hume argued that moral responsibility requires a necessary connection between one's character and one's actions. Without such a connection, he argued, the action would not say anything about the agent and could not be attributed to him. Hume thought so because he considered the idea of necessary connection as the crucial element of our idea of causation. In other words, he thought that without a necessary connection there would be no causal connection between the agent and his action.[73] Hume's followers thus concluded that concluded that far from being incompatible with determinism, free will relevant for moral responsibility requires determinism or rather necessity (in some psychological form at least).

However, this idea is not popular any more for several reasons. First, and the most important one is that most philosophers nowadays reject the claim that causation requires a necessary connection between the cause and its effect. Holding the opposite view is now regarded as very unscientific because of the widely accepted interpretation of quantum mechanics, according to which processes at the quantum level, the level of the smallest particles do not obey deterministic laws (and no one wants to deny that at that level there is no causation). Another important reason why compatibilists don't find this 'Humean' route to compatibilism appealing anymore is that they don't like the idea that our responsibility depends on empirical discoveries about the basic structure of our universe. Finally, the idea

[72]Hobart thus says that "'absence of determination, if and so far as it exists, is no gain to freedom, but sheer loss of it; no advantage to the moral life, but blank subtraction from it" – quote taken from Neil Levy's book *Hard Luck: How Luck Undermines Moral Responsibility*, 41.

[73]Chisolm also seems to accept this claim (as well as Ayer). The idea that control is impossible without causation together with the idea that there is no indeterministic causation seems to be Chisolm's main motive for defending agent-causation. See Roderick Chisholm, "Human Freedom and the Self," in *Free Will*, ed. Robert Kane (Blackwell, 2001)

that free will requires psychological determinism is very unpopular today. For, most philosophers think not just that we have no evidence for the existence of some sort of psychological laws, but they even argue that it makes no sense to postulate such laws.[74]

For these reasons, most contemporary compatibilists argue for the so called 'even if compatibilism', which says that we can be morally responsible even if causal determinism is true. Compatibilists of this sort don't think that indeterminism is inhospitable to free will and moral responsibility (some even argue that it is hospitable), but only that it is irrelevant. In their view, if we cannot be free and morally responsible in a deterministic world, we cannot be free and responsible in an indeterministic world either. Their challenge to libertarians is thus to show how indeterminism *contributes* to control, that is, to show how it provides space for control that is impossible to have if determinism is true.[75]

Thus, every libertarian theory faces two challenges. First, they must show that indeterminism does not eliminate or diminish control that may be available if determinism is true. Second, they must explain how indeterminism 'helps,' that is, they must explain how indeterminism contributes to control relevant for moral responsibility. I will argue in this chapter that arguments to the effect that indeterminism undermines control, although initially very plausible, don't speak decisively against the possibility of libertarian free will. However, I will also argue that we have very good reasons to think that that indeterminism is irrelevant for free will and moral responsibility.

[74] The most influential argument against psychological determinism is due to Donald Davidson. See Donald Davidson, "Psychology as Philosophy," in *Philosophy of Psychology*, ed. S. Brown (Harper and Row, 1974).
[75] The objection that indeterminism undermines free will can nowadays usually be found in the texts of the so called skeptics about moral responsibility – philosophers who think that moral responsibility is impossible. See, for instance, Neil Levy, *Hard Luck: How Luck Undermines Moral Responsibility*, (Oxford: Oxford University Press, 2011).

3.2 Types of Libertarian Theories

In order to understand properly the challenges to libertarianism related to indeterminism, it is necessary to say something about the types of libertarianism. But, in order to do that we must understand the positive aspect of libertarian theories – the account of agency required for libertarian free will. The only way to do that is to consider different conceptions of agency in general, because different types of libertarian agency correspond to those conceptions. So, we need to see what the possible answers to the questions 'what is action' are.

There are two most general answers to this question: causal and non-causal. The first is that that an action is an event caused in a certain way. The second is that it is an event of a certain kind, that is, it is an event with certain intrinsic features.

Philosophers who find the first answer plausible, the so called causalists about action, usually find compelling one part of an old argument against libertarianism which says that an event which is not caused is random or occurs by chance and that no one has control over events that are random and/or occurs by chance. In their view, doing and controlling are essentially causal phenomena. Moreover, they think that doing something is bringing about and bringing about is causing. They disagree, however, about what kind of entity a cause of an action must be. This is an important question because an event that constitutes some action can have a cause but fail to be an action. For instance, an arm can *rise* as a result of an electric impulse from an external source, without being *raised* by a person whose arm it is. For the latter to be the case, according to some causalists – the 'event-causalists'- the arm's rising must be caused by some agent-involving events in the appropriate way. In particular they think that it must be caused by the (onsets of) agent's beliefs and desires and/or his intention to raise it. Other causalists – 'agent-causalists' - think that an event that constitutes

action must be caused by the agent either directly or indirectly via other events caused directly by the agent as a substance. But both groups of causalists believe that action can be reduced to causally related entities which are not in themselves active.

Non-causalists have a problem with this last claim. They reject the view that action can be reduced to a relation of entities that are essentially non-active. In their view, actions are basic elements of reality which can only be described by pointing to their essential features. Thus, we can explain what they are by pointing to their specific phenomenology, intentionality, spontaneity etc, but we cannot explain them reductively. For, in their view, the assumption that causality is essential to agency leads to some difficult problems. For instance, they argue that if we consider acting as causing we must postulate another event which is the agent's causing of the event which constitutes his action and that leads to an infinite regress, because explaining the activity of that new event requires postulation of a new causal relation etc.

Now, it is important to mention that theories of action just mentioned have different metaphysical implication. Among those theories, the event-causal theory is considered as the least metaphysically problematic. That is so because on a prevailing view of causation all causation is causation of events by events. In addition, this theory does not imply anything about the truth of determinism. It is compatible both with the truth and the falsity of determinism. Similar is true about the non-causal theory. Although non-causalists see actions as special kind of events, and argue that something can be action even if it has no cause, they don't require the absence of causation or even absence deterministic causation. That is not true of agent-causal theory, because most agent-causalists argue that agent causation requires not just the absence of deterministic causation in the causal history of an action, but absence of any sort of event causation.

Finally, it is important to say something about the phenomenon of acting for a reason. This is important because virtually all libertarians agree that free will is a power that can be exercised for a reason. In addition, this is important because one of the main reasons why some philosophers think that indeterminism undermines control is that it eliminates the possibility of a certain kind of reasons-explaining of action. So, what is acting for a reason?

Again, there is a causal and a non-causal account of this phenomenon. According to causal account, a person acts for a reason when her action or event that constitutes her action is caused by her reasons (understood as the belief/desire pairs) or states representing reasons (understood as states of affairs). Philosophers who accept this account argue for it by claiming that acting for a reason implies the existence of reason explanation of why the action occurred, which in their view, cannot exist unless the reason which explains the action is not what actually moved the agent to action. And, this can be true only if the reason caused the action.

The non-causalists, on the other hand, explain the phenomenon of acting for a reason in terms of the action's directedness toward some goal reflected in the agent's mental states at the time of action. They believe that this is the only way to understand this phenomenon because only in that way can we understand the tight connection between reasons for which the agent acts and the action done for those reasons. In their view, the problem with the causal account is that it cannot capture this fact; and it cannot do that because causation is essentially blind (anything can cause anything). In their view, this is reflected in the problem of deviant causal chains - the difficulty to differentiate causal chains which constitute acting for a reason and causal chains which do not constitute that phenomenon.

Let us now turn to the relation between these conceptions of agency to libertarian accounts of free action. As I have mentioned, libertarians' conceptions of free agency correspond to their conceptions of agency in general. Thus, for libertarians who accept event-

causal theory of action free action as of an event caused by the agent's mental events but not determined by those or any other events. Libertarians who accept agent-causal theory of action usually conceive free action as an event caused by the agent (as a substance), and usually claim that this entails that the agent or the agent's causing of that event is not caused by any other event or substance. Finally, libertarians who are non-causalists about action think of free action as of an undetermined event, with certain intrinsic features, which (the event) may or may not be (in fact) caused.

However, libertarians' conceptions of action and free action do not always overlap. For instance, some libertarians think that event-causal theory is good as an account of action but not as an account of free action.[76] They think that free action requires a metaphysically more demanding conception of action such as agent-causation. On the other hand, some libertarians think that unlike action, free action requires a combination of causation by the agent and causation by agent-involving events. This is the so called integrationist account.[77]

Interestingly, the mismatch in those libertarians' conceptions of agency and free agency is mainly a result of their concerns about control. Agent-causal libertarians usually think that event-causal theory of action cannot explain how the agent can have sufficient control to be morally responsible his or her action. Recently, however, some philosophers have argued that a theory of action that a libertarian holds does not make a difference to their position with respect to the worries about control. That is they have argued that the worries about control are simply the result of the libertarian requirement that the action must be undetermined by previous events. I will argue that these philosophers are right. But, more importantly, I will argue that although the arguments of those who think that indeterminism undermines control have strong appeal, those arguments are not conclusive.

[76] Randolf Clark, for instance, hold this view. See Randolph Clark, *Libertarian Accounts of Free Will*, (Oxford: Oxford University Press, 2003): 93-116.
[77] This is exactly Clark's view. See Randolf Clark, *Libertarian Accounts of Free Will*, 133-148.

But, before presenting these arguments I must say that there are two things that I will simply assume in the following discussion. First, I will assume that only indeterminism at the moment of action is the relevant kind of indeterminism from the libertarian perspective. This is in conflict with the view of the so called *deliberative libertarians* who think that it is enough for free will if the agent's deliberation is indeterministic in the sense that it is not determined which thoughts come to mind in the process of deliberation. Deliberative libertarians believe that indeterminism so located gives the agent independence from his environment without diminishing his or her control in action. This is so, they explain, because agents anyway don't have (direct) control over the thought that come to their mind in deliberation. However, this proposal is clearly unsatisfactory, in my view, because the sort of indeterminism in question obviously does not provide more control to the agent than he or she would have if in a deterministic world.

Second thing that I will assume is that concerns about control can be divorced from concerns about agency. That is, I will assume that arguments that follow don't have a goal to show that a specific account of action is problematic because it is not a good account of agency but because it is not a good account of free agency.

3.2 The Problem of Control

At the beginning of this chapter I presented an argument against libertarian free will based on the claim that indeterminism implies randomness or chance. That argument can be developed in various ways depending on how one wants to defend the claim just mentioned. As we have seen, one way of supporting this claim is to say that indeterminism is incompatible with causation. As I have mentioned, this way of developing the basic argument is not popular any more. Very similar way of supporting the crucial premise is by arguing

that indeterminism is incompatible with acting, i.e. that there cannot be indeterministic actions. But this way of arguing against libertarianism is not very promising either because none of the three main theories of action considered above requires that an event be determined in order to count as action. Yet another way of developing the basic argument against libertarianism uses the inference rule Beta which is the main element in the Consequence Argument for incompatibilism. This version of the argument is supposed to show that no one has a choice about an event that is undetermined. Van Inwagen calls this argument the "third strand of the Mind Argument," but I will call it the No-Choice Argument. In the following section I will argue that this argument fails to show that indeterminism undermines control.

3.2.1 The No Choice Argument

The argument which is the topic of this section is focused on the event-causal libertarianism. It aims at a conclusion that if event-causal theory of action is correct no one has a choice about any undetermined event, including one's own actions, by showing that the agent in the following story did not have a choice about the action he performed:

> Let us consider the case of a hardened thief who, as our story begins, is in the act of lifting the lid of the poor-box in a little country church.' 9 He sneers and curses when he sees what a pathetically small sum it contains. Still, business is business: he reaches for the money. Suddenly there flashes before his mind's eye a picture of the face of his dying mother and he remembers the promise he made to her by her death bed always to be honest and upright. This is not the first occasion on which he has had such a vision while performing some mean act of theft, but he has always disregarded it. This time, however, he does *not* disregard it. Instead, he thinks the matter over carefully and decides not to take the money. Acting on this decision, he leaves the church empty-handed.[78]

[78] Peter van Inwagen, *An Essay on Free Will*, 127-128.

At first sight the thief's decision not to take the money seems like a paradigm example of free and morally responsible action, especially if we assume that it was not determined. But, according to van Inwagen, every incompatibilist must deny this. For, as van Inwagen argues, the principle Beta which leads to the conclusion that no one has a choice about an action that is determined leads us to the conclusion that the thief had no choice in this situation for his decision (if it was indeterministically caused by the mental states which explain why he decided not to take the money, e.g. his desire to keep the promise he gave to his dying mother and belief that not taking the money is a means to satisfying that desire). Representing the thief's desire/belief complex with DB, his decision not to take the money with R, and his lack of choice with N, van Inwagen argues for this claim in the following way:

> (1) The thief's repentance was caused but not determined by DB, and nothing besides DB was causally relevant to the thief's repentance [assumption for conditional proof]
>
> (2) N DB occurred [premise]
>
> (3) If (1) is true, then N(DB occurred → the thief repented) [premise]
>
> (4) No one (including, of course, the thief) had any choice about whether the thief repented.
>
> (5) If the thief's repentance was caused but not determined by DB, and nothing besides DB was causally relevant to the thief's repentance, then the thief had no choice about whether he repented.[79]

[79] Ibid., 147.

As van Inwagen points out, although this argument, if sound, shows that free will is incompatible with indeterminism, it is of no use to the compatibilist because it is valid only if Beta is a valid rule of inference and so only if compatibilism is false. The argument is acceptable only to those who think that free will is impossible because it is incompatible with both determinism and indeterminism. To show that free will is possible, a compatibilist must reject Beta. On the other hand, the incompatibilist seems to have no choice but to reject premise 3 of the argument. For, as van Inwagen explains, the first premise is simply an assumption for the purpose of the argument, and the second premise is obviously true. The latter is the case because given the description of the case the thief could have a choice about the occurrence of DB only indirectly by making some earlier choice. But, since the same argument can be applied to every (earlier) choice this strategy would ultimately lead to the agent's first choice which is based on DB which could not be result of any earlier choice. Thus, we can simply assume that the thief did not have a choice about DB.

But, on what grounds can the incompatibilist reject premise 3? Van Inwagen admits that he does not have an answer to this question if event-causal theory of action is correct. He admits that it is puzzling in that case that this premise should be false. That is, he thinks that it is not clear how the thief could "have a choice about whether R follows DB if DB is insufficient for R, and nothing else is even causally relevant, save negatively, to R." As he observes, a possible solution to this problem would be to introduce agent-causation, that is, to assume that the thief as a substance caused R. For, in that case it would be false that nothing else besides DB was causally relevant to the occurrence of R. However, van Inwagen does not find this solution satisfactory because the notion of agent-causation appears to him "more puzzling than the problem it is supposed to be a solution to."[80] So, he concludes that this argument represents a serious challenge to libertarianism. The only good news for the

[80] Ibid., 151.

libertarian, in his view, is that the argument represents a bigger problem for the compatibilist than for the libertarian, because, in his view, the rejection of *Beta* seems more implausible than the rejection of the third premise of this argument.

However, in my opinion, there are two serious problems with this argument. First, it is not clear that there is a valid version of the principle *Beta* which this argument could use. For, as the discussion of the first chapter has showed, there are only two versions of the principle that can survive counterexamples: the one which involves the necessary connection between the antecedent and the consequent in the second premise of the Consequence Argument (Warfield and Finch's version of Beta), and the one restricted to deterministic scenarios (Warfield and Crisp's version). Now, Dana Nelkin has shown that the No-Choice argument works equally well with the former version of the principle.[81] However, I am not aware of any attempt to show that it works with the latter version, and it seems that no one could show that because the principle is applies only to deterministic cases.

But, even if there was a version of Beta that would render the argument valid, I am not sure that the argument would work, because I am suspicious about the premise 3 of the argument. The premise certainly sounds plausible. However, if we follow van Inwagen in assuming that the thief decided not to take the money and that it was open to him not to do that, it is not clear why we should believe that he had no choice. Perhaps, as Randolf Clark suggests, we should think so because the thief in this case lacked "freedom-level control." But, unless Clark wants to suggest that freedom level control consists in ability to choose to choose, which leads to the infinite regress, I see no reason to conclude that the thief had no choice, except that it was undetermined. But if indeterminism is the only reason why we should think that the thief lacked choice, No-Choice argument begs a question.

[81]See Dana Nelkin, "The Consequence Argument and the Mind Argument," in *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*, ed. Paul Russel and Oisin Deery, (Oxford University Press, 2013), 126-134.

Perhaps, however, the premise 3 can be supported by pointing out that the thief was not able to *ensure* what he will decide by making an evaluative judgment or deciding what he should do. Some philosophers have argued that indeterminism diminishes control exactly for that reason.[82]

However, as Clark points out, this argument is applicable only to instances of indirect control, control that is exerted over an event by performing an action, and not direct control, which is not exercised by performing some other action, which is the topic of this discussion. In addition, as Robert Kane points out, "it does not follow that because you cannot determine which of a set of outcomes occurs *before* it occurs, you lack control over which of them occurs, *when* it occurs."[83]

Thus, I conclude that if the No-Choice Argument shows something it shows that event-causal theory is not an adequate account of action. Otherwise, it fails as an argument against event-causal libertarianism.

3.2.2 The Luck Argument

The so called 'luck argument' is probably the most popular argument against libertarianism. It goes roughly like this: If right up to the moment of occurrence of an action, which is of some significance for the agent, there was a chance that it would not occur, its occurrence was partly a matter of luck (good or bad). And to the extent that it was a matter of luck that the action would occur, the action was not under the agent's control, because luck entails the absence of control. But, since responsibility requires control, to the extent that an action is a result of luck, to that extent the agent lacks responsibility for that action.

[82] According to Randolph Clark, Alfred Mele has argued for this position. See Randolph Clark, *Libertarian Accounts of Free Will*, (Oxford: Oxford University Press, 2003), 74-77.
[83] Robert Kane, *The Significance of Free Will*, 144.

Of course, philosophers who advance this argument don't claim that an action must be entirely a matter of luck if indeterminism is involved in its production. They recognize that it may not be a matter of luck that the agent will perform *some* action and that the agent will perform a certain *type* of action. What is a matter of luck is that the agent performed that particular action *rather* than some other action that it was open to the agent to perform. And this is a matter of lack precisely because of indeterminism involved in choice between those actions.

To support this claim, the proponents of the luck argument usually tell a story of two agents who are identical in all relevant respects right up to the moment of decision. The moment of choice is the moment when their stories start to diverge. One agent makes a good decision while the other makes a bad decision. The reader is then asked to consider what it is about the two agents that accounts for the difference in their decisions, i.e. the fact that of them made a good decision and the other made a bad decision. And the answer seems to be: nothing. But if that is the case, they conclude, its occurrence is just a result of chance or luck.

Libertarians take this argument very seriously. Some of them think that the only way to avoid its conclusion is to accept the claim that free will requires special power such as agent-causal power which cannot exist in worlds in which all events are caused by prior events. For, they believe that the worry that indeterminism implies luck is essentially related to the assumption that all there is to free will is causation by certain events. But the critics of libertarianism argue that this is not so. On the other hand, event-causal libertarians think that the postulation of special powers does not eliminate the worry about luck, but instead introduces new problems related to the possibility of powers in question. Furthermore, they believe that the worry about luck can be eliminated in some other way.

The most sophisticated and the most criticized event-causal libertarian reply to the problem of luck has been offered by Robert Kane. I discuss his reply in the following section.

3.2.2.1 Kane's Event-Causal Response to the Luck Argument

According to Robert Kane, one of the reasons why it seems that indeterminism undermines free will is the idea that an act which is not determined must occur accidentally, or must be arbitrary, capricious or something like that. He illustrates this by mentioning Schopenhauer who ridiculed libertarian freedom by comparing it to a freedom of a man whose legs suddenly indeterministically started to move, although he did not plan to move them. In addition, he presents a case of a person who after a thorough deliberation about whether to go for a vacation to Hawaii or Colorado concluded that all things considered Hawaii is a better, but then suddenly indeterministically opted for Colorado. In these cases, as Kane observes, we are inclined to say that the persons' actions were a result of chance.[84]

However, as he points out, indeterminism does not necessarily have this effect. In particular, when the agents do what they intend to do and what they have reasons to do or what they want to do, the fact that there was a chance that they would not do it (or simply that they would do otherwise) does not imply that the action was arbitrary, capricious, random or accidental. Kane illustrates this point with an example of a husband who tried to break a glass table top and succeeded although there was a chance that the table top would not break. Also he mentions Austin's example of a sniper shooter who managed to hit his target in spite of a chance that he would fail.

According to Kane, the fact that indeterminism does not undermine our responsibility for actions that we are inclined to do is the reason why many libertarians have been attracted

[84] See Robert Kane, "Responsibility, Indeterminism and Frankfurt-style Cases: A Reply to Mele and Robb," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, edited by David Widerker and Michael McKenna (Ashgate, 2006), 193-214.

by Leibniz's claim that motives incline without necessitating. However, as he observes, this idea cannot save the libertarian freedom from the objection that indeterminism undermines control. For the libertarian freedom requires that at least for some actions right before their performance it was open to the agent to freely perform an alternative action or to freely refrain from performing them. But, if the agent's motives always inclined the agent toward the action he actually performs, for that very reason, his performing of an alternative action would be arbitrary, capricious and irrational, and so not free. Thus, according to Kane, for libertarian freedom to be possible, there must be some occasions in a person's life when the person has more than one option from which to choose which is from her perspective rational. In addition, it must be the case that whichever of those options the agent would choose, he or she would do it as a result of his or her will. And, finally, the agent must have voluntary control over which option he or she chooses, or which action he or she performs. Kane calls these conditions the plurality conditions for free will.

According to Kane, when an agent performs an action under these circumstances and the action is undetermined, he or she has the power not only to do what he or she wills to do, but also to determine or to shape one's own will. For that reason the agent is not only responsible but also ultimately responsible for an action performed in these circumstances. Kane says that an agent is ultimate responsible for an actions when that action has no sufficient condition for which the agent is also not, at least partly, responsible. And this is not only true about the actions which satisfy these plurality conditions but also about some actions which do not satisfy them, (the actions which are not undetermined or in which the agent does not have an alternative which has a motive to perform) if the will from which he or she performs them partly originates and is partly formed by some action which does satisfy them. For obvious reason, Kane calls actions which satisfy the plurality conditions the self-forming action or SFA's.

Of course, Kane is aware that by postulating these conditions he does not eliminate all reasons for thinking that indeterminism undermines free will. In particular, he is aware that someone might still say that even if an agent decided for a reason and because he wanted to do it and it was open to the agent to do otherwise, he was not responsible because what he did was a matter of luck. He says that he feels the pull of this intuition but he believes that it can be eliminated by looking more closely at circumstances in which SFA's occur. He observes that SFA's occur when persons have equally strong motives to perform actions of different kinds, actions which for them have incommensurable values. In a situation of this kind the agent is not indifferent to what he will choose, because for each option he has reasons for choosing that it *rather* than some other one option. (Kane agrees with the critics of libertarianism who argue that being indifferent to one's options or having the liberty of indifference, cannot contribute to one's freedom.) But because of the equal strength of one's motives one experiences an inner conflict and must invest effort to make a choice. The effort corresponds to an indeterministic process in the agent's brain and when the agent makes a choice, the choice results indeterministicaly from that effort. Finally, according to Kane, if in a situation of moral conflict the agent decides to perform the action that morality recommends his decision is a result of the agent's effort, and if he decides to do what he is tempted to do that is the result of his not allowing his effort to succeed.

Now Kane says several things about the nature of effort in an attempt to eliminate the residual worry that it is a matter of luck which decision follows the effort. He says, for instance, that indeterminism is not a problem here because it is the result of one's own will, it comes from inside so to say. In addition, he says that it is not the case that the agent first makes effort and then indeterminism or chance resolves the conflict. It is rather the case that indeterminism and effort are fused. Moreover, according to Kane, in SFA's indeterminism of a decision is a result of the effort's *indeterminacy*. And since the effort is indeterminate, it is

not possible to say that the agent would, in the same situation, perform a different action on another occasion or that the identical agent would do otherwise in a different possible world. Finally, Kane suggests that in SFA's the agent's effort is not only directed toward one option, that is, the agent is not only trying to make one decision. Instead, the agent is trying to make several incompatible decisions, and for that reason, whichever option he settles with, in the end, it will be the option he was 'aiming at' all along. Obviously, this last suggestion is supposed to eliminate the worry about luck regarding SFA's by making them similar in crucial respect with the actions of the husband or the assassin in the above mentioned examples.

However, as critics have pointed out, there are at least two problems with Kane's answer to the luck argument.[85] First, it is not clear whether it makes sense to talk about efforts to perform two incompatible actions. What is more, it seems that irrationality involved in such an attempt would rather diminish than enhance one's control. The second problem is that if an action is free because the effort in which it originated, the effort must be a free action or an SFA. But in that case we must have an account of how an effort can be a free action. However, Kane has not offered such a theory.[86] Thus, it seems that he has not given a satisfactory answer to the challenge posed by the luck argument.

Nevertheless, I think that Kane's discussion of the problem of luck contains some important insights about the relation of indeterminism and free will. In particular, his

[85]See Clark, Randolph. *Libertarian Accounts of Free Will*, (Oxford: Oxford University Press, 2003), 82-92.

[86]According to Neil Levy, Kane's dual efforts strategy also has a bizarre consequence that the agent is responsible for what he has not done to the same extent to which he is responsible for what he has done. That is so, in his view, because the agent's responsibility for what he has done is grounded in his effort to do it and he has invested the same effort in doing otherwise. Levy says the following: "If Kane's account of responsibility is correct, responsibility is doubled: if we are responsible for our directly free actions, then we are also and equally responsible for the counterfactual actions we also tried to perform in the same circumstances. Dual control and dual rationality leads to dual responsibility: responsibility for what we do, and for what we would have done instead. But if I am equally responsible, either way, then what I actually do does not matter, at least so far as my responsibility is concerned. I deserve neither praise for my right actions, nor blame for my bad—at least neither to the exclusion of the other. Perhaps I deserve both, and in equal measure." Neil Levy, *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*, 63.

observation that indeterminism does not undermine responsibility in cases when the agent's reasons strongly favor the action the he actually performs sounds very convincing. But, let us see whether the agent-causal libertarian has a better reply to the luck argument.

3.2.2.2 Agent-Causal Libertarianism and the Luck Argument

An important element in Kane's theory that I haven't mentioned above is the Free Agency Principle. This principle says that the libertarian should not postulate entities that are not also required for theories that do not require indeterminism, i.e. the entities or relations which could not in principle exist in worlds that are deterministic. All the above mentioned conditions satisfy this criterion. But perhaps giving a convincing reply to the luck argument requires rejecting this principle. Most agent-causal libertarians reject the Free agency Principle. They think that to have full control over one's behavior one must be able to cause one's actions directly as an enduring substance and not via the states and events involving it and for many this power cannot exist in a deterministic world. One of the most prominent contemporary agent-causal libertarians, Timothy O'Connor, explains how exactly the idea of agent-causation eliminates the problem of luck in the following passage:

> Given the presence of desires and intentions of varying strength, making certain outcomes more likely than others, the agent possesses no further power to determine which outcome in fact is brought about. The determination is a product of the propensities of the agent's states, and the agent doesn't seem to directly control which propensity will 'fire.' If we imagine two identical agents in identical circumstances, with one agent nondeterministically choosing alternative A and the other choosing B, it seems a matter of luck from the standpoint of the agents themselves which alternative occurs in which person.
> Supposing there a power of agent causation has the virtue that it seems to avoid this 'problem of luck' facing other indeterminist accounts. Agent

> causation is precisely the power to directly determine which of several causal possibilities is realized on a given occasion.[87]

I don't see, however, how the concept of agent-causation could help libertarians to avoid the problem of luck. This concept helps if we assume that agents whose activities involve only causation by their mental states are just passive observes of what goes on within them. In that case, to say that someone is an agent-cause is simply to say that he is an agent. But the luck objection seems to go beyond the question of activeness and concerns the agents' ability to control how they exercise their agential powers (i.e. how they exercise their control). For, the main assumption behind the objection is that there is not enough control if there is no *explanation* of the contrastive fact that it was exercise in this rather in some other causally open way. For, as Alfred Mele puts it, "if nothing accounts for the difference, the difference is just a matter of luck."[88] But, given this interpretation of the luck objection, it is difficult to see how the concept of agent-causation helps to eliminate the problem because there seem to be no explanation either of the contrastive fact that on a given occasion the agent exercised his agent-causal power in one way rather than some other way.

But is the contrastive explanation really so important? I am not sure clear what the right answer to this question is. But one thing is clear, if the requirement that a free action must have a contrastive explanation simply amounts to the requirement that a free action must be determined by the agent's antecedent mental states, the luck argument against libertarianism simply begs the question.

Thus, although the agent-causal libertarianism does not seem to provide a better protection from the luck argument, if the luck argument rests only on the assumption that free action must have a contrastive explanation, the argument is inconclusive. But the requirement

[87] Timothy O'Connor, "Agent-Causal Power," in *The Philosophy of Free Will*, ed. Paul Russell and Oisin Deery. 243.
[88] Alfred Mele, *Free Will and Luck*, (Oxford: Oxford University Press, 2006), 59.

for contrastive explanation is not the only reason why many philosophers have been reluctant to embrace libertarian conception of free will. A more powerful reason against libertarianism has been that the agent-causal libertarianism, which has traditionally been regarded as a more promising libertarian account, cannot account at all for the rationality of the agent-causal activity. The worry is that the agent causal contribution in the production of the agent's activity amounts to something like a blind shot a reflex or a totally inexplicable occurrence. Since this worry can be traced back to Leibniz, I call it the Leibnizian objection to agent-causal libertarianism.

3.2.2.3 The Leibnizian Objection to Agent-Causal Libertarianism

A picture that naturally comes to mind when we think about what goes on in the agent when he exercises his agent-causal power is the picture of an inner arena in which passions and reason as separate agents fight for the sympathies of the agent and try to convince him to follow their advice. The agent is free when he has the power to choose whether he will listen to reason or passions and determine his will independently of their influences. This seems to be the picture of libertarian free will that Leibniz had in mind when he compared the libertarian free will with a queen

> seated on her throne, whose minister of state is the understanding, while the passions are her courtiers or favorite ladies, who by their influence often prevail over the counsel of her ministers. One will have it that the understanding speaks only at this queen's order; that she can vacillate between the arguments of the minister and the suggestion of the favourites, even rejecting both, making them keep silence or speak, and giving them audience or not as seems good to her.[89]

[89]G. W. Leibniz, *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*, Trans. E. M. Huggard (La Salle, Ill.: Open Court, [1710] 1985), 421.

This picture of the role of will or the agent-causal power is no doubt appealing if acting for reasons is understood as a causal phenomenon. On that understanding of acting for reasons, the agent causal activity which is exercised for a reason must have some reason (desire or belief) in its causal history. But, if the will is a source of activity totally independent of passions and the reason, this is not possible unless the will has its own reasons (separate from the reasons provided by the agent's reason and his passions). But, then, the same picture appears concerning the relation between those new reasons and the will. That is, we must postulate some other will and some other reasons on the basis of which she chooses between those other reasons. Therefore, on pain of infinite regress we must settle with the view that the agent-causal activity is not done for a reason at all. But, since free will must be a power to act for reasons (it would not be of much value if it could only be exercised irrationally) the agent-causal power could not be free will.

A natural response to this challenge by agent-causal libertarians is to deny the existence the detachment of agent from his reasons. The will is not a separate agent inside of the agent who decides what the agent will decide. There is only one agent who makes decisions on the basis of his or her own reasons.

But it is not clear whether it is possible to make sense of this picture if we stick to the idea that acting for a reason should be understood causally. Randolf Clark thinks otherwise. He argues that his integrated agent-causal account according to which acting from free will consists in simultaneous causation of action by the agent and his reasons solves the problem. He explains why he thinks so in the following response to Galen Strawson:

> An integrated agent-causal account, then, is crucially different from the Lebnizian view. On the latter, as Strawson sees it, the agent "exercises some special power of decision or choice" (1986:53). However, an agent-causal account does not attribute to free agents any special power of *decision* or *choice*. Rather, it attributes to them a *causal* power that is distinct from the causal powers that can be exerted by events involving them. The agent does

> not *decide* which decision to make, and she need not decide which reasons to make effective; she *causes* a certain decision or other action, one that is made or performed only if it is caused by certain reasons.[90]

But is it enough for agent causal power to be exercised for a reason that the agent and his reasons simultaneously cause his action? And does this theory really provides the answer to the above presented problem that the agent-causal libertarian needs? More precisely, does it manage to avoid the Leibnizian objection while at the same time showing that agent-causal power gives more control to the agent over his actions than he would otherwise have, if all the causation involved were the causation of his actions by his reasons. It is not clear that it can do so because, as Neil Levy observes, agent-causal libertarians see free agents as "difference-makers," and Clark's theory leaves no room for that idea. Levy elaborates on this objection in the following passage:

> It is when the agent's pre-existing reasons run out—when she has reflected carefully, in the light of her preferences, desires, beliefs, goals, and values, and seen how things stand with her options—that the agent-causal power must be called upon to exert its final push. At this point in proceedings, however, we cannot cite the agent's pre-existing reasons as her reason for the final agent-causal push, understood as a difference-maker, on pain of double counting.[91]

In my view, this consideration speaks decisively against the sort of agent-causal view which includes the claim that reasons must have some kind of causal role in acting for a reason. However, the agent-causal may have other options. In particular, the agent-causal libertarian may not have to endorse the causal theory of acting for a reason. In that case, there is no reason to think that agent-causal activity cannot be exercised for some reason. Consequently, there is no reason to think that the exercise of the agent-causal power on particular occasion must be a matter of luck.

[90] Clark, 176.
[91] Neil Levy, *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*, 69.

So, we have seen that there are essentially two routs to the conclusion that indeterminism does not undermine control. One of them is not very useful for libertarians but it might be useful for the theorists who do not require robust alternative possibilities for moral responsibility. I have in mind Kane's observation that indeterminism does not undermine responsibility for actions that we intend to perform and have decisive reasons to perform. The other route consists in showing that there is no reason to worry that indeterminism implies absence of control apart from the worry that an action that is not determined cannot be done for a reason, and in showing that there is reason for this latter worry only when a specific version of agent-causal libertarianism is concerned.

However, although libertarians in general don't have to worry that indeterminism undermines control, they have to worry that powers which their theories ascribe to free agents don't provide more control to them than do powers that compatibilist theories ascribe. That is, they should worry that indeterminism does not have a special value in so far as acting freely is concerned. I will argue for this claim in what follows.

3.3 The Problem of Value

The problem of value of indeterminism for free will is obvious in the case of event-causal libertarianism. Libertarian theories of this sort do not postulate any special positive powers in comparison with compatibilist theories. The only difference between these theories and compatibilist theories is the requirement of indeterminism. Thus, event-causal libertarian free will differs from compatibilist free will only by an absence. And it is not clear how a mere absence of something can provide agents with control necessary for moral responsibility and other things for which free will is valued such as autonomy, self-determination etc. Using the symbol C for conditions required by the compatibilist theories, and referring to a world in

which these conditions are satisfied as a C-world, Gary Watson presents this objection to what he calls soft-libertarian theories[92] in the following passage:

> The basic incompatibilist intuition is something like this: determinism is inconsistent with the existence of certain human capacities and powers, say autonomy or self-determination, that are central to the meaning and dignity of human life. The knowledge that something is a C-world is not enough to determine whether or not the individuals in this world enjoy this possibility or are doomed to utter impotence and emptiness. What is incredible is to suppose that these values are secured by the mere truth that some of the relevant events and processes are indeterminate. If C is not enough to ground those values, introducing the negative condition of indeterminacy will not do it either.[93]

So, event-causal libertarianism and other libertarian theories which consider indeterminism as a condition logically independent from other conditions for free will cannot account for the special value that indeterminism has for libertarian free will.[94] For, according to Watson, for indeterminism to have a meaningful role in a libertarian theory, it must be required for the satisfaction of the positive conditions that the libertarian theory postulates. This requirement is satisfied by some agent-causal libertarian theories, according to which the agent-causal power can exist only if determinism is false. So, perhaps these theories avoid the objection that indeterminism has no relevance for free will?

There are at least two reasons to think that the answer to this question is 'no.' First is related to the fact that it is not clear how the agent-causal power provides the agent with enhanced control over his or her behavior. Agent-causation clearly provides a person with more control compared to the person who does not possess that power if agent-causation is necessary for agency. But if both the person who is an agent-cause and person who is not an

[92] Watson uses the label 'soft-libertarianism' for all libertarian theories which don't require forms of agency that could not exist in a deterministic world.
[93] Gary Watson, "Soft Libertarianism, Hard Compatibilism," in *Agency and Answerability* (Oxford: Clarendon Press, 2004), 203.
[94] As far as I know these other theories include non-causal libertarian theories because the positive aspect of these theories, the non-causal theory of action, does not say that agency requires indeterminism.

agent cause acted and it was open to them to do otherwise and there was no explanation why either of them acted in the way they acted rather than in the alternative way opened to them, it is not clear why we should think that the agent-cause had more control over his action. It seems true, though, that assuming that agent-causing cannot be caused by prior events, agent-causal power gives agents more independence from their environment then the event-causal power. But again, this independence consists only in an absence and we are not told how this absence enhances control except that it provides space for the power which is intrinsically the power to control. Therefore, even if this power is conceptually possible and even if it provides enhanced control (which for all I know might be the case), the account which requires this power is a bit mysterious. And it is not surprising that it has been accused of being just a label for what the libertarian needs.

The second reason to be suspicious that agent-causal libertarianism provides more control than libertarianism of the event-causal type is that it is not clear that agent-causation may not exist in a deterministic world. In my view, the strongest reason to think that there may be deterministic agent-causation is that agent-causation is a species of object-causation and that it is possible that all causation is of that type. Dana Nelkin calls this view of causation the Kantian view because some indications that Kant understood causal relata in this way. On the Kantian view, objects cause changes in other objects in virtue of their natures and their circumstances which together may determine the object's exercise of its causal power. But if it is true that other objects as objects have the power to produce certain effects in spite of being determined to 'do' that by their natures, why couldn't the same be true of agents? Agents, no doubt, have a special nature, which includes powers to understand reasons and to act on the basis of reasons, but they could also cause their actions in virtue of those powers and they may be determined in their exercise.

As Nelkin observes, (if I understand her correctly) this picture of causation does not apply to free agents only if they don't act in virtue of their natures, that is, if their behavior is inexplicable by their reasons, passions, habits etc. This is the case, she assumes, with the so called non-propensity libertarian views. To this category seem to belong the agent-causal views according to which the agent's reasons do not influence their actions causally. According to Nelkin, the problem with the non-propensity views is that it is not clear how they provide enhanced control (compared to compatibilist views). Nelkin's reason for this claim seems to match my first reason for the suspicion in the value of agent-causal free will. For she says that these accounts are "in key respect negative: the agent causes, but *not* in accordance with any propensities that parallel the propensities of her reasons."[95][96]

Finally, there is an empirical objection to all agent-causal accounts raised by Derk Pereboom. According to Pereboom, since the exercise of agent-causal power cannot be explained by the agent's states at the moment of action (or immediately before the action) it is then a coincidence that the agent's actual choices match the propensities that those states impose. In addition, it is an unbelievable coincidence that agents always choosing to perform actions that the changes in their brains cause them to perform.

3.4 Conclusion

In conclusion, there is no evidence that every libertarian view fails either because indeterminism undermines control or because it cannot provide more control than it is

[95] Dana Nelkin, *Making Sense of Freedom and Responsibility*, 92.

[96]But, I think that even the views which don't assign causal role to the agents reasons or more generally views according to which the agent's reasons do not determine their propensities toward certain types of behavior face the challenge of explaining why such agent-causing cannot be determined. For, in certain situations when agents have strong reasons for a certain action and there is no interference with the agent's abilities (the circumstances are normal), it is inconceivable that the agent will not act in a certain way or refrain from doing something. Thus, most ordinary sane people would not torture innocent people for small amount of money. It is not clear why they cannot be agent-causes of their action is such situations.

possible to have in a deterministic world. Some libertarian views do seem to face these problems though. The event-causal libertarianism and other sorts of 'soft-libertarianisms' indeed provide no more control than similar views that do not require indeterminism. On the other hand, some forms of agent-causal libertarianism seem to fail because they cannot explain how the agent-causal activity can be guided by reasons. But, an agent-causal view which does not face these problems is conceivable. It is the view according to which the agent's causing of their actions is not explicable in terms of the agent's states and which do not act for reasons because their reasons cause their actions. However, although this view seems coherent, it is not very informative. For it is not clear how the power it refers to confers the agent more control than they could have in a deterministic world. Besides, like all forms of agent-causal libertarianism the view is quite mysterious because it postulates the existence of wild coincidences in the world in addition to other heavy metaphysical assumptions.

Therefore, libertarianism is a theory that indeed rests on "obscure and panicky metaphysics" as Peter Strawson famously remarked. This, together with the conclusion of the first chapter that determinism is incompatible with the ability to do otherwise is a good reason to try to explain the nature of free will in terms of some property or process which does not presuppose the ability to do otherwise and the falsity of determinism. Many philosophers suggest that such property or a process can be found by focusing on the human ability to act for reasons. In the next chapter I will explore this suggestion.

CHAPTER 4: SUSAN WOLF'S REASON VIEW

As we have seen in the second chapter, philosophers are divided concerning the question about the relevance of ability to do otherwise for moral responsibility. Many philosophers think that the ability to do otherwise is necessary for moral responsibility, and many think that it is not. However, there are also some philosophers who think that moral responsibility only sometimes requires this ability. In other words, some philosophers regard free will necessary for moral responsibility as asymmetric. A view of this kind has been introduced in the contemporary debate on moral responsibility by Susan Wolf. According to Wolf, free will is asymmetric in the sense that it involves the ability to do otherwise when one acts wrongly or for wrong reasons, but not when one does the right things on the basis of the right reasons. This claim is sometimes called the Asymmetry Thesis, or simply Asymmetry.

Interestingly, asymmetry is an implication of the main claim of Wolf's theory that it is necessary and sufficient for moral responsibility that a person has the ability to recognize the right reasons and act in accordance with and on the basis of them. The Asymmetry falls out of this condition because when one recognizes and acts in accordance with and on the basis of the right reasons, the condition is satisfied (because acting for the right reasons implies that one can act for the right reasons) and there is no reason to ask whether the person could have done otherwise; but when one fails to exercise those abilities, ability to do otherwise is important because doing otherwise in those circumstances is doing the right thing for the right reasons.

The Asymmetry thesis is not very popular among philosophers. In fact, only one philosophers beside Wolf - Dana Nelkin - seem to endorse it.[97] However, I argue in this chapter that the number of its supporters does not reflect its plausibility. I argue for this claim by arguing that Wolf's view explains our intuitions about epistemic and freedom conditions for moral responsibility, both when it comes to responsibility for bad actions and responsibility for good actions. Following Wolf, I call her view the Reason View. I argue for the plausibility of this view when it comes to its explanation of moral responsibility for bad actions by comparing it with other views which consider rational abilities of some sort essential for moral responsibility, but which don't require ability to do otherwise (I call these views, including the Reason View, the 'rationalist' views). I do that in the first part of this chapter. In the second part, I argue that the Reason View provides the correct account of (and has the right implications when it comes to) moral responsibility for good actions. I do that by comparing the Reason View with what I call the 'traditional view', according to which free will and moral responsibility always require ability to do otherwise. I argue that proponents of the traditional view make mistake in considering ability to do otherwise as more fundamental when it comes to free will and moral responsibility than the ability to do the right thing for the right reasons.

4.1 The Rational for Asymmetry

The idea that moral responsibility can be understood solely in terms of rational capacities is very popular among philosophers. Moreover, many philosophers agree with Wolf that moral responsibility requires the ability to recognize and act in accordance and on

[97] Perhaps also Michael Smith, Philip Pettit accept Wolf's view, but I cannot say that with confidence now. See, Michael Smith and Philip Pettit, "Freedom in Belief and Desire," in *Mind, Morality and Explanation*, Oxford: Oxford University Press (2004): 375-396.

the basis of the right reasons. For that reason it is an interesting phenomenon that so few 'rationalists' accept the asymmetry thesis which seems to be an obvious consequence of this understanding of rational capacities in question.

Clearly, the only possible explanation of this phenomenon is that most philosophers have a different understanding of the relevant than Wolf and her followers. Consider first Wolf's claim that in order to be morally responsible one must be able to recognize the right reasons. All philosophers endorse this condition to the extent that moral responsibility requires the ability to grasp the relevant facts. For all philosophers agree that people with certain forms of cognitive impairments are not morally responsible for their actions. However, there is no universal agreement about the extent to which one's cognitive capacities must be developed and what the facts that morally responsible agents must be able to grasp are. In particular, it is a matter of debate to what extent it is important to have a correct view not just of non-normative facts (e.g. whether one is stepping on someone's foot or on a piece of wood) or non-moral facts but also of normative or moral facts (e.g. whether it is good or bad to step on peoples' feet). Most philosophers agree that moral responsibility requires ability to grasp both kinds of facts. But, according to Wolf, it is not enough just to have *some* grasp of these facts to be morally responsible. For instance, in her view, it is not enough to have the concepts of (morally) good and bad and consider some things good and some bad to satisfy the epistemic requirement on moral responsibility. In her view, for that purpose, one must also be able to know what is *truly* good and what is *truly* bad. For this reason Wolf describes the epistemic faculty required for moral responsibility as the ability to recognize the True and the Good. (This is the reason why she calls her view the Reason view with the capital R.)[98]

[98] I am not sure, though, that it is right to characterize Wolf's requirement that the responsible agent must be able to recognize the right reasons as the epistemic condition for moral responsibility although this requirement involves essentially cognitive capacities.

This explains the difference between Wolf's view and another "rationalist" view – the so called Real Self Views. According to a version of this view, proposed by Gary Watson, an agent has free will if he has the power to translate his values into action, that is, if he can act in accordance with and on the basis of his conception of what is good. This view is asymmetric when it comes to free will because it implies that in order to act with free will one does not have to be able to act against one's values, but only in accordance with them. This is not the case when it comes to moral responsibility. For, according to this view, the agent is responsible only when he actually uses this ability (because only in that case his actions express his real self). However, the asymmetries of these views (regarding free will) are different because on the Real Self View, ability to do otherwise might be required for good actions as well as for bad actions given that this view does not say what one must value in order to have free will. Thus, according to this view, if one values doing bad things (i.e. has a mistaken conception of what is good), one needs the ability to act badly in order to exercise free will in doing something good (which is not the case according to the Reason View). The Reason View's demanding epistemic requirement on moral responsibility is thus one of the reasons why it entails Asymmetry.[99]

However, this is not sufficient to explain how the Reason View grounds Asymmetry. It is also necessary to see how proponents of this view understand the freedom requirement for moral responsibility, that is, how they understand the ability to *govern* one's actions in accordance with and on the basis of the right reasons. For, it is possible to accept Wolf's claim that in order to be morally responsible one has to be able to grasp what is truly good without accepting Asymmetry. This is so because the ability to act in accordance with and on the basis of the right reasons can be understood either as a general or as a specific ability. According to the standard interpretation of this distinction, general abilities are abilities that

[99] I am not sure though that we have here only a difference in the epistemic requirement for moral responsibility.

one can have even on occasions when one cannot exercise them due to the lack of opportunity to exercise them.[100] Specific abilities are abilities that one has only if, in addition to the general abilities, one has the opportunity to exercise them. Thus, one might have a general ability to play the piano even when there are no pianos around (Nelkin's example), but lack the specific ability to play it in that situation because of the lack of opportunity to play it. So, if the ability to act in accordance with and on the basis of the right reasons is understood as a *general* ability, the agent might be responsible for his action even if he is *not able to exercise that ability* on a given occasion.[101][102] This interpretation of the relevant rational ability has been suggested by R.J. Wallace.

Similar way of dealing (away) with Asymmetry is available to the proponents of the so called 'reasons-responsiveness' views that do not ground responsibility in the rational abilities of *agents* but in the specific characteristics of the *mechanisms* on which they act. Thus, according to the most prominent view of this kind, Fischer and Ravizza's reasons-responsiveness theory, an agent is morally responsible for his action if his action results from the activity of (his own) moderately reasons responsive mechanism, which is such that it sometimes results in actions which the agent has sufficient reason to perform. What is crucial here is that just like the general ability mentioned above, one can have the ability that is central to this view even if one cannot exercise it in the particular circumstances. That leaves

[100] See Dana Kay Nelkin, *Making Sense of Freedom and Responsibility* (Oxford: Oxford University Press), 67.
[101] As Ferenc Huoranszki points out, this understanding of the distinction between general and specific abilities is problematic. For, if having the specific ability implies having the ability to exercise some general ability there is a danger of infinite regress because the ability to exercise some ability would also seem to require ability for its own exercise and so on ad infinitum. In addition, it is not clear "why would the 'power to exercise a general ability' be any more specific or less general than the ability which is exercised." See Ferenc Huoranszki, Freedom of the Will: A Conditional Analysis (New York: Routledge, 2011), 25.
4. In addition to this requirement, Fischer and Ravizza's postulate the requirement of ownership which says that the agent has to have taken responsibility for the mechanism which leads to his actions. See John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press 1998).
But, this is not essential here.

space to the proponents of this theory to say that an agent can be morally responsible even if he was not able to do otherwise on a given occasion.

In sum, Asymmetry is the result of a specific understanding of the epistemic and freedom requirements for moral responsibility. That is, it is a consequence of the claim that moral responsibility requires the ability to recognize the *right* reasons (the True and the Good) and to respond to those reasons *in the specific circumstances,* (i.e.the specific ability as opposed to a general ability to govern one's behavior in accordance with reasons). This explains why not all views that regard rationality as in some sense essential for moral responsibility entail Asymmetry.

But is the way of interpreting rationality which entails Asymmetry the right way of interpreting it? In other words, is the Reason View preferable to other 'rationalist' accounts of moral responsibility?[103]

In what follows I argue that it is by pointing to its advantages to the rival accounts. In particular, I argue that this view fits better our ordinary understanding of moral responsibility and that it is more efficient in dealing with some philosophical problems.

4.1.1 Reason View and Real Self View(s)

[103]In one respect rationalist theories which do not entail Asymmetry are certainly more appealing than the Reason View. For, according to these theories, metaphysical questions about the fundamental structure of the world are irrelevant to the question of moral responsibility. In particular, according to the above mentioned rival views to the Reason View, whether anyone is free and morally responsible for anything does not depend on whether the laws of nature are deterministic or indeterministic. That is not the case with the Reason View because it requires ability to do otherwise for responsibility for bad actions. So, the Reason View does not rule out the possibility that no one has free will and acts responsibly if determinism is true. This is no doubt an advantage of the 'symmetric rationalist' accounts of free will. For, if they are correct our moral responsibility for bad actions 'does not hang on a thread,' so to say.[103] But, this advantage of these theories concerns only their consequences, and I am not interested here in the consequences of the views under examination but only in whether the views capture correctly our intuitions about the conditions of moral responsibility.

As I indicated above, the Real Self Views say that agents are morally responsible for what they do if their actions express their real selves. As I also mentioned, Watson identifies one's real self with one's values.[104] Harry Frankfurt, another famous representative of this view, identifies the real self with the agent's higher-order volitions which he wholeheartedly identifies with. Higher-order volitions are agent's (higher-order) desires that certain desires move him to action. According to Frankfurt, an agent is morally responsible when the desire that actually moved him to action matches the higher-order volition which he wholeheartedly identifies with.

The appeal of these theories rests on the fact that it is hard to deny that an agent is morally responsible for an action if the action is truly attributable to him and it is hard to deny that the latter is the case if the action expresses his true self. And, it is not implausible to think that there is such a thing as real self. For, some sources of motivation seem more closely related to oneself than some other sources of motivation. In addition, these theories can account for a wide range of our intuitions about responsibility in particular cases. For instance, they can explain why we don't hold people responsible for actions that are the results of direct influences of external factors such as electrical stimulations of brain or forces acting on their bodies. According to these theories, people are not responsible for actions of these types because they do not originate in their real selves. Also, they can explain why we don't consider little kids or animals morally responsible. This is the case, according to the Real-Self Views, because these beings do not have true selves, either because they cannot have them or because they don't have them yet. Finally, these theories seem to do a good job in explaining why factors such as phobias, addictions, and manias seem to undermine responsibility. For the intuition that agents whose behavior is determined by these factors are not responsible can be explained by pointing out that these factors, although in an obvious

[104] I am not sure though that Watson uses the phrase 'real-self.' I think the phrase is due to Susan Wolf.

sense internal, are in the *relevant* sense alien to the agent. They are alien to the extent that they are not part of the agent's real self.

However, these views cannot explain all of our intuitions about moral responsibility. In particular, they cannot explain why we don't find morally responsible agents who act from values or desires which they wholeheartedly identify with, but which they have acquired via some sort of brainwashing. In addition, it is also not clear whether they can account for the intuition that psychopaths are not morally responsible for their actions. Finally, for obvious reasons, these views do not allow for morally responsible action against one's better judgment or akrasia.

The Reason View has obvious advantages over the Real-Self Views in these respects. For, beside the fact that it can explain all the intuitions that the Real-Self View can explain, it can explain our intuitions about the brainwashed agents. According to the Reason view, those agents are not morally responsible because they lack the ability to recognize the right reasons or to form the right values. In that respect, the agents in question are similar to little children, animals and psychopaths and dissimilar to agents who suffer from phobias, addictions, and manias that are not responsible because they lack the ability to *act* for the right reasons.

Perhaps the problem with the Real Self Views just mentioned can be solved in a way which does not require different understanding of rational capacities necessary for moral responsibility. Perhaps it can be solved by adding a historical condition of moral responsibility. That is, it might be solved by requiring that the agent must acquire the relevant values or desires in a particular way, or that he must not acquire them in some way (e.g. via brainwashing). This might be so although as I will show in the next chapter this suggestion brings new difficult problems related to compatibility of free will and determinism. What is

important here, however, is that if we focus only on the properties of the agent at the time of action, the Reason View explains much better our intuitions than the Real-Self View.[105]

What about the reasons-responsiveness view of Fischer and Ravizza?

4.1.2 Reason View and Reason-Responsiveness View

According Fischer and Ravizza, an agent is morally responsible for an action if he exercises guidance control in performing it.[106] Guidance control is a kind of control that one might have even if one lacks the ability to do otherwise and it is the opposite of the regulative control – control which involves ability to do otherwise. According to Fischer and Ravizza, a person has guidance control when his actions are produced by his own deliberative mechanism which is moderately responsive to reasons. Moderate reasons-responsiveness concerns the state of the agent or his decision-making mechanism at the time of action. It has two aspects: receptivity and reactivity to reasons. A mechanism has these properties if the agent who acts on it (thanks to the mechanism) regularly recognizes and sometimes reacts to the right reasons. (Reasons here obviously do not refer to the states of the agent but to states of affairs.)[107] This condition enables Fischer and Ravizza's theory to explain all the intuitions that the Real Self is able to explain, but also to explain why brainwashed agents are not responsible. For, these agents do not recognize the right reasons at the time of action i.e. the mechanism from which they act is not receptive to (the right) reasons. In addition, the (moderate) reasons-responsiveness condition enables Fischer and Ravizza's to make a distinction between the agents who act against their better judgments, the weak-willed agents,

[105] Besides, the Real-Self Views are known as 'non-historical' views. Here I just wanted to point out that there is a different way to respond to the difficulties that the views of this type encounter.
[106] For the most detailed presentation of Fischer and Ravizza's account of guidance control, see their *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press 1998).
[107] Reason-responsiveness is usually explained in terms of what happens in other possible worlds.

and compulsive agents. The difference, on their view, consists in the fact that the former but not the latter act on a mechanism which is *reactive* to reasons. This means that the mechanism which leads the former to act weakly in the actual world or the actual situation, does not lead them to act weakly in some other world. In other words, the compulsive agent, unlike the weak agent, is such that he would never perform a different action that he has sufficient reason to perform, as long as the mechanism on which he acts remains the same.[108]

Thus, the reason-responsivness condition seems to enable Fischer and Ravizza to draw all the distinctions between different actions regarding moral responsibility that want to make. However, as Dana Nelkin notices, there are two serious worries about their account. First, it is unnatural to assume that responsibility depends on some properties of the *mechanisms* on which the agents act, even if those mechanisms are just the ordinary psychological processes. It is much more natural to think of responsibility as a function of some properties of the *agent*. Furthermore, if the former were the case, it seems that we should blame a mechanism and not the person for some wrongdoing because the person might have been simply unlucky to have a vicious mechanism, or it might have such a mechanism because she was a victim of some sort of manipulation. The second condition of guidance control, the ownership condition, is meant to eliminate this sort of objection. It is also meant to eliminate the objection of incompatibilists who argue that one cannot be responsible for an action that was causally determined by factors outside of one's control. But, whether or not this condition succeeds in eliminating these further worries, it remains true that the Reason view tracks better our intuitions about responsibility by grounding it in the properties of agents rather than in some mechanisms internal to the agent.

[108] Several critics have pointed to difficulties concerning the issue of identification of mechanisms on which agents act. For an interesting presentation of these difficulties see Gary Watson, "Reasons and Responsibility," in *Agency and Answerability: Selected Essays* (Oxford: Clarendon Press 2004), 294-301.

However, a much more serious problem for Fischer and Ravizza's view is that they ground responsibility in *general* powers or dispositions of the agent's mechanism to respond to reasons. For it is not clear how an agent can be responsive to reasons in the sense relevant for moral responsibility if he (or his mechanism) could not respond to the reasons on the particular occasion. Fischer and Ravizza believe that Frankfurt-style examples eliminate this worry. But, since I reject those examples relevance, this seems to be a pretty serious worry to me. And, since the Reason View does not face this problem, in my eyes this represents an important advantage of the Reason View over the Reasons-Responsiveness view.[109]

I turn now to Wallace's view. His view is similar to Fischer and Ravizza's view in that it requires general ability rather than specific ability to act in accordance with what the agent considers to be the right reasons. Wallace, however, unlike Fischer and Ravizza, ascribes the relevant ability to the agent rather than to the mechanism that leads him to action. In addition, Wallace's belief that moral responsibility does not require ability to do otherwise (the specific ability to act for the right reasons) does not rest solely on Frankfurt-style examples.[110] He has also an alternative strategy for showing that, which I consider in what follows.

4.1.3 Reason View and Wallace's View

According to J. R. Wallace, ability to do otherwise is relevant for moral responsibility only if its lack makes blaming unfair. And since it is not fair to blame someone for something only if the person has an excuse or is exempt from responsibility, Wallace undertakes a comprehensive analysis of actual and possible excuses and exemptions to see if they can be

[109] For a more detailed presentation of these objections see Nelkin, *Making Sense of Freedom and Responsibility*, 18-20.
[110] Perhaps Fischer and Ravizza have some additional arguments in favor of their view that I have not considered here. If that is so, I must admit that my discussion of here is incomplete.

understood without reference to (the lack of) ability to do otherwise. According to Wallace, the result of his analysis is positive: the full range of actual and possible excusing conditions can be explained solely by reference to the conditions that show that the agent has done nothing wrong, while the full range of exempting conditions can be explained by reference to the conditions that show that the agent did not have the *general* abilities to recognize moral reasons and regulate behavior by their light.

Let me clarify this by focusing first on Wallace's account of excuses.[111] According to Wallace, valid excuses show that the agent hasn't done anything wrong.[112] That does not mean, he explains, that they show that nothing bad has happened (that the agent has not produced some bad state of affairs), because, in that case, there would be no need for an excuse.[113] It means rather that the wrongness of an action depends essentially on the quality of the agent's will, that is, it depends on the agent's intentions in performing the action. Thus, one has a valid excuse if one has not brought about the bad result intentionally, that is, if his action does not reflect bad quality of his will. On that account, for instance, inadvertence excuses because when someone does something inadvertently, the action does not reflect the quality of one's will, or more simply, it does not express one's choice.[114] Thus, Wallace

[111] Wallace explains what the aim of his discussion of excuses is in the following way: "it must be established that all of our considered judgments of excuse can be explained without appealing to a principle of fairness that supports incompatibilism, such as the principle of alternate possibilities. To show that this is the case, I need to identify an alternative principle of fairness that explains why people do not deserve to be held to blame when they have not violated the moral demands to which we hold them. I then need to consider the various excuses in sufficient detail to establish that all of our judgments of excuse can be accounted for in terms of this alternative principle; and finally, I need to show that this principle would not support the incompatibilist's conclusion that determinism is a kind of generalized excuse." R. Jay. Wallace, *Responsibility and the Moral Sentiments* (Harvard University Press, 1998), 127.

[112] Wallace in this respect follows Peter Strawson. See Wallace, *Responsibility and the Moral Sentiments*, 126.

[113] The distinction between justification and excuse is very interesting in this context. Wallace discusses this distinction on pages 120 and 121 in his book.

[114] Why is the quality of choice so important? Wallace explains this by saying that "one can be said to have complied with a moral obligation only when there is present a relevant quality of choice. Someone who inadvertently bumps into me, thereby knocking me out of harm's way, has in no sense complied with the obligation of mutual aid; by contrast, a person can be said to have *complied* with the obligation if she acted out of a choice to save me from harm--even if the choice was based on reasons of a self-interested rather than a moral nature. Similarly, one cannot be said to have *violated* a moral obligation in the absence of a relevant quality of choice." Wallace, *Responsibility and the Moral Sentiments*, 142.

seems to show that at least some excuses can be understood without reference to alternative possibilities. According to Wallace, alternative possibilities *appear* to be required for moral responsibility only because some excusing factors such as compulsion or coercion eliminate them. However, in his view, the reason why these factors excuse is not the fact that agents in these circumstances cannot do otherwise, but the fact that in those circumstances agents do not do bad things intentionally. For, following Frankfurt, Wallace argues, that a factor which eliminates alternative possibilities undermines responsibility only if it explains the agent's actual behavior.[115]

Wallace further argues that when some excuses are concerned it does not even appear that the lack of ability to do otherwise is the reason why the agent is not morally responsible.

> Thus if I harm someone inadvertently, the natural way to beg for excuse would be to say not "I couldn't help it," but rather "I didn't mean to hurt you." Similarly, a coercive threat of torture if I do not open the safe would not ordinarily be thought to prevent me from taking some other course of action, but only to make such alternatives extremely unattractive.[116]

Now, concerning exemptions, Wallace says that they are factors in virtue of which people (either temporarily or permanently) fail to be appropriate objects of moral appraisal, i.e. the kind of beings to which moral appraisal properly applies.[117] Is the lack of specific ability to do otherwise – the ability to do otherwise in the circumstances - such a factor? According to Wallace, the answer is negative because exemptions apply only to agents who lack general abilities, rather than to agents that cannot exercise them in particular circumstances. More

[115] Wallace also says that if "one decides not to meet one's obligation to rendezvous with the friend, but one discovers that one has all along been handcuffed to one's chair or locked in one's office, these forms of constraint will not be valid excuses for the failure to meet the obligation. In these cases, despite the presence of physical constraints, one's omission nevertheless expresses precisely the kind of choice that our moral obligations prohibit." Ibid., 142.

[116] Ibid., 151.

[117] Wallace also says that "exemptions are unlike excuses in being less localized: whereas excuses block responsibility for particular acts an agent has performed, exemptions make it inappropriate to hold the agent accountable more generally. Ibid., 154.

precisely, according to Wallace, exemptions concern only general *rational* abilities. That is so, in his view, because it is unreasonable to hold someone responsible for violating a moral obligation if one could not grasp the reasons that support that obligation and regulate his behavior by their light ("and thereby to avoid the sanctions associated with failure to meet the demand"[118]). For, according to Wallace, not being able to exercise a general ability in the particular circumstances does not count as an exemption. For instance, a person tied to a chair is not blameworthy for not trying to save a drowning child, not because of the lack of ability to jump into the water and swim, but because she did nothing wrong. Finally, according to Wallace, certain factors such a hypnosis or brain manipulation do exempt by (temporarily) incapacitating, but they do that by eliminating general abilities to recognize and apply reasons, rather than by eliminating specific abilities to do that.

However, in my view, Wallace's arguments for the claim that his account of excuses and exemptions is superior to the account which cites the lack of ability to do otherwise are not convincing. This is so because the plausibility of his arguments depends on his interpretation of the distinction between general and specific abilities, which is in my view completely mistaken. The main problem with his account of this distinction is the identification of the specific abilities with *abilities to exercise* general abilities, which is, as Ferenc Huoranszki points out, very problematic because it leads to infinite regress. For if in order to exercise some ability, person needs some other ability to exercise that ability, it is natural to ask whether the person has the ability to exercise that latter ability as well. So, the fact that Wallace misidentifies specific abilities explains why they seem irrelevant for him. For, it is difficult to see how anyone can be deprived of something that does not exist and thereby excused for something or exempt from responsibility.[119]

[118] Ibid., 162.
[119] I need to say a bit more here.

In addition, Wallace' observation that the lack of ability to do otherwise does not account for some excuse does not seem correct. It is true that inadvertence excuses because it entails the lack of intention to cause harm. But it is also true that when we hear that someone did something inadvertently we want to know whether he could have avoided doing what he did in that way. For instance, we want to know if he could have paid more attention to what he was doing or if he could have prevented being distracted etc.

Finally, Wallace's account and other accounts which deny the relevance of alternative possibilities for all actions are in conflict with the principles that 'ought' implies 'can' and that wrongness of an action implies that one ought not to perform it. Thus, an agent is blameworthy only if the agent did something wrong. And, an action is wrong only if the agent ought not to do it. But, according to the 'ought implies can' principle, the agent ought not to do something, only if he can avoid doing it, that is, if he can do otherwise. So, it seems that Wallace's account and similar accounts require rejection of the 'ought implies can' principle, which seems to be a very unpalatable consequence of those views.[120]

Interestingly, however, proponents of the Reason View do not face this problem. For, as Dana Nelkin points out, the 'ought implies can' principle does not require ability to do otherwise for praiseworthy actions. That is so because to be praiseworthy one must do the right thing which together with the 'ought implies can' principle implies that to be praiseworthy one must be *able* to do the right thing. But, since one's doing the right thing shows that one is able to do the right thing, the 'ought implies can principle' does not entail

[120] John Martin Fischer rejects this principle on the basis of Frankfurt-style examples. See John Martin Fischer, "'Ought-Implies-Can', Causal Determinism and Moral Responsibility," *Analysis* 63 (Jul., 2003): 244-250. However, as I have argued in the second chapter, these examples do not show that ability to do otherwise is irrelevant for moral responsibility.

that in order to be praiseworthy one must be able to do otherwise. Thus, this highly esteemed ethical principle seems to support asymmetry of the Reason View.[121]

With this I conclude this admittedly very brief discussion of the 'rationalist' alternatives to the Reason View. I believe, nevertheless, that I have offered decisive reasons for preferring the Reason View over the other 'rationalist' accounts of moral responsibility. The main advantage of the Reason View over those views is that it implies correctly that responsibility for bad actions requires ability to do otherwise.

But to see whether the Reason View really represents the correct account of moral responsibility, it remains to examine whether it has the right implication when it comes to responsibility for good actions. In other words, we need to check the claim that moral responsibility for good actions performed for good reasons does not require ability to do otherwise. This is a very interesting topic because, on the face of it, the Reason View does not agree with the common sense in this respect. However, I will argue in what follows that our intuitions don't speak against this aspect of the Reason View when we look more closely at particular cases, that is, when we consider what being able to do otherwise means in particular cases.

4.2 Ability to do otherwise and responsibility for the right actions

In chapter 2 I argued against one of the most popular challenges to the claim that free and responsible action requires ability to do otherwise – the challenge based on Frankfurt-style examples. However, that challenge is not the only reason why some philosophers find the connection between ability to do otherwise and moral responsibility problematic. Another

[121] For a detailed discussion of the relation between moral responsibility and the 'ought implies can' principle, see Dana Kay Nelkin, "A Rational for the Rational Abilities View: Praise, Blame, and the Ought-Implies-Can Principle," in *Making Sense of Freedom and Responsibility* (Oxford: Oxford University Press, 2011), 98-116.

set of cases which are much less fanciful and perhaps quite common in real life represents for some philosophers a more serious reason for doubts about that connection. The most famous case in this set is considered to be Martin Luther's rejection of the proposal to recant his criticism of the Catholic Church. For this reason, the cases in question are sometimes called the 'Martin Luther cases.' These cases represent agents who are doing something that, by their own lights, they have decisive reason to do, and seem praiseworthy what they are doing, although they lack the ability to do otherwise. Martin Luther, thus, by his own testimony, could not do otherwise (he reportedly said: Here I stand, I can do no other), but, this inability, intuitively, did not undermine his moral responsibility for his action. Moreover, it seems that this inability made him even more responsible and praiseworthy for what he did (or rather for what he refused to do). For, the inability in question was not a result of external forces or blind inner urges, which intuitively undermine responsibility, but of his conviction that doing otherwise would be wrong.

Another instance of this type of case is this one provided by Susan Wolf: while standing on a beach woman sees a child drowning. Seeing clearly what she should do, without hesitation she rushes into the water to save the child. According to Wolf, if in fact the woman could not do otherwise, that would not be a reason not to consider her praiseworthy. For, again, the reason why she could not do otherwise was not some factor that intuitively undermines responsibility but her clear understanding of what is the right thing to do.

These examples obviously put to question the claim that there is a necessary connection between the concept of alternative possibilities or ability to do otherwise and the concept of morally responsible agency. In that respect they are similar to Frankfurt-style examples. However, unlike Frankfurt-style examples, these examples seem to show only that there is no such connection when agents act on the basis of their recognition of what the *right* thing to do is. For there are no cases of this type with agents performing bad actions. In other

words, there are no cases of this type with agents who are blameworthy although he could not avoid performing actions for which they are blameworthy. Therefore, the Luther cases do not support the claim that moral responsibility does not require ability to do otherwise in general, but only when doing otherwise is irrational, that is, they support the asymmetry suggested by the Reason View.

Of course, Luther cases do not prove that the ability to do otherwise is not fundamental for free will and responsibility. For, there may be good reasons to think that agents in those examples in fact have the ability to do otherwise or that they are not morally responsible. For instance, it might be that the fact that no external force compelled Martin Luther and the woman on the beach to act in the way they did (together with other details of the case) suffices for saying that they had the ability to do otherwise. Alternatively, it could be that we consider them morally responsible because we apply to them an insufficiently robust notion of responsibility. For instance, it might be that our judgment that they are praiseworthy in these cases simply expresses our positive evaluation of their characters. Or it might be that we considered them responsible only in a *derivative* sense because we see their actions as results of their earlier directly or non-derivatively free actions (which they had the ability not to perform). Finally, there may reasons to think that moral responsibility requires ability to do otherwise which override our intuitions about the Luther cases.

I will argue, however, that none of these responses to Luther cases ultimate succeed. I will focus on the question about the ability to do otherwise of agents in these cases. For, I don't see a reason to think that agents in these cases lack direct responsibility for their actions except the assumption that direct responsibility requires ability to do otherwise, which is exactly the assumption these cases put to question. Besides, the suggestion that the agents in these cases are only derivatively responsible sounds implausible because there is a clear difference between these cases and paradigm cases of derivative responsibility (e.g. a case of

a person who caused a car accident in a state of drunkenness and who intentionally got in that state).[122]

4.2.1 Van Inwagen's Argument

A plausible way to describe Luther cases is to say that they are situations in which agents find alternative actions indefensible and have no desire or have very weak desire to perform them.[123] But if this is the correct description of those cases, an argument by Peter van Inwagen, very similar to his modal argument for incompatibilism, seems to show that the agents in those cases cannot do otherwise. The argument goes roughly like this:

1) It is unavoidable for a person S at a certain moment t that S at t regards act A as an indefensible act, and has no desire or a very weak desire to perform A, and has no way of getting further relevant information about A.
2) It is unavoidable for S at t that (if S at t regards act A as indefensible, and has no desire or a very weak desire to perform A, and has no way of getting further relevant information about A, S is not going to do A immediately after t).

Hence,

3) It is unavoidable for S at t that S will not do A immediately after t.[124]

[122] For the suggestion that Luther cases belong to cases of derivative responsibility see Robert Kane, *The Significance of Free Will* (Oxford: Oxford University Press 1998), 77-79. For an interesting criticism of this suggestion see Ferenc Huoranszki, *Freedom of the Will: A Conditional Analysis*, 166-175. For an interesting discussion of views that Luther cases represent cases of direct responsibility see: Gary Watson, "Volitional Necessities," in *Responsibility and Answerability* (Oxford: Clarendon Press, 2004), 88-122.

[123] I am not sure if everyone would agree with me about this. But, I don't see a problem in simply stipulating that this is what these cases are.

[124] See Peter van Inwagen, "When is the Will Free," *Philosophical Perspectives* 3 (1989): 409.

This argument is based on a version of the inference principle Beta (the key element in van Inwagen's modal argument for incompatibilism), which van Inwagen calls Beta prime. Beta prime is simply a version of Beta restricted to the particular agent and a particular time. As van Inwagen shows, Beta prime can be derived from Beta.[125] Thus, according to van Inwagen, every incompatibilist (about ability to do otherwise and determinism), at least, should conclude that this argument is valid.

But compatibilists may say that this argument is a sort of reductio ad absurdum of incompatibilism. For, they may argue that since we obviously act freely and choose in situations of this sort, which implies that we can do otherwise in those situations, the incompatibilist reasoning must be invalid (assuming that the premises are true). They may support this claim with the standard compatibilist observation that the fact that agents in these situations never *would* do otherwise does not entail that they *could* not do otherwise.[126][127]

I don't find this compatibilist reply convincing because I don't think that we can simply assume that free action and ability to make a choices requires ability to do otherwise, especially in the light of the existence of examples presented above. However, I am not convinced by van Inwagen's argument either because I am not sure that its first premise is true. I think that the second premise is true because I agree with van Inwagen that if we add all the information that we implicitly accept about the agents in such situations (e.g. that the agent will not unexpectedly go berserk), it becomes inconceivable that in those circumstances the agent will do otherwise. To see this, imagine yourself in such circumstances. For instance,

[125] Ibid. 410.

[126] For a reply of this sort to van Inwagen see: Ferenc Huoranszki, *Freedom of the Will: A Conditional Analysis* (New York: Routledge, 2011), 155-158.

[127] Moreover, they may use these cases as evidence that compatibilism is not mysterious. For, the fact that agents always act in certain ways in certain situations because of certain reasons, shows that the agent's reasons may explain their actions without necessitating them. Or as Leibniz famously observed, reasons incline but do not necessitate. In other words, it is not the case that they simply mysteriously always choose as the laws of nature say they will choose.

imagine that someone offered you a thousand dollars to torture an innocent person?[128] Is there a coherent scenario in which you would accept that offer given your present state of mind and other background facts?

But, van Inwagen's first premise is problematic. He supports it by saying that like most of our beliefs and desires our belief that some course of action is indefensible is something that we just find ourselves with. He also says the following in its support:

> If you offered me a large sum of money, or if you promised and I believed you could deliver-the abolition of war, if only I were to change my attitude toward A, I should not be able to take you up on this offer, however much I might want to. It is barely conceivable that I have the ability to change my attitude toward A over some considerable stretch of time, but we're not talking about some considerable stretch of time; we're talking about right now.[129]

However, it is not clear to me that what is said here supports van Inwagen's first premise. Of course, I cannot make it the case that I don't have *now* some belief that I now have. For, I cannot make contradictions true. I cannot even make it the case that I have a different belief immediately after this moment. But, even if I could that, it would not make a difference to the argument because we are interested in whether I could have a different belief *now*. So, the claim about unavoidability of my belief now plausibly concerns my past ability to influence my present beliefs and desires. And, although I cannot change my attitudes at will, it is not clear that my attitudes are something that I *just* find myself with. For, I believe that I can influence my future attitudes by what I do now and that I have done that in the past with respect with my current attitudes. In that case, it may not be true that my present beliefs and desires are simply unavoidable for me.[130]

[128] This thought experiment is due to Daniel Dennett. See Daniel C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Oxford: Clarendon Press, 1984), 133-135.
[129] Van Inwagen, "When is the Will Free," 408.
[130] It is interesting to observe how the 'unavoidability' of laws of nature and the distant past differs from the 'unavoidability' of our present attitudes.

Therefore, this argument does not *prove* that agents in Luther cases cannot do otherwise. There may be some other argument which does prove that, but I must admit that I am not aware of any such argument. Alternatively, there may be a way to fix this argument but I don't see at this moment how that could be done.[131] So, I am not sure that the incompatibilist, or anyone else for that matter, must hold that agents in Luther cannot do otherwise. Nevertheless, I find the claim that it is *possible* that agents in Luther cases cannot do otherwise very plausible. I will explain why I think so by discussing the value of ability to do otherwise in the cases in which we actually perform good actions. For, in my view, consideration of the value of having the ability to do otherwise in such circumstances shows that moral responsibility for good actions does not require ability to do otherwise.

4.2.2 The Value of Ability to Do Otherwise

The phrase 'ability to do otherwise' is usually understood in a very general way in the free will debate. It is simply understood as the ability to perform an alternative action or to omit performing the actually performed one. The discussion is then usually focused on the question of its compatibility with determinism and indeterminism. For these purposes the quality of the alternative action seems irrelevant. Whether it is good or bad, rational or irrational is of no importance to the discussion. However, according to the proponents of the Reason View, these further characterizations are crucial for determining the value of the ability to do otherwise, that is, for determining whether this ability is relevant for moral

[131] Perhaps the problem with the argument can be fixed by applying the unavoidability operator to some interval of time instead to particular moments. In that case, it would make sense to say that the agent could not change his attitude in that interval of time. However, this might provide the agent with the freedom to choose the exact time of performing the action which his attitudes support. O'Connor says that it is even in this situation open to the agent to wait a bit longer with making a decision. Alternatively, it might be that the relevant sense of unavoidability simply concerns my inability to create contradictions (to make it the case that I don't have now some attitude that I have now). See Timothy O'Connor, *Persons and Causes*,101-107.

responsibility in the first place. The reason why they think so emerges when we consider situations in which we don't have sufficient reason to do otherwise or in which doing otherwise has a negative value. For, in such cases the ability to do otherwise seems to be an ability that no rational agent would want to exercise.[132] And it makes sense to ask the questions that Susan Wolf asks in the following passage:

> Why should one want an ability that one never wants to exercise? Why should one care about being locked in a room—or better, in a world—out of which one cannot *conceivably* want to go? Why should one mind if, to put it in extreme terms, one is *inescapably* sane?[133]

It is important to notice that in this passage Wolf distinguishes the question about the value of exercise of some ability from the question about the value of *having* that ability (although she obviously expresses doubt about the relevance of this distinction). In other words, Wolf recognizes the *possibility* that having the ability to choose a less optimal option has value even if its exercise is worthless. This is very important because the value of having the ability to do otherwise is clearly more fundamental in this context than the value of its exercise. For, it is the having of the ability to do otherwise and not its exercise that it supposed to ground our control over our actions and make what we do up to us.

[132]Susan Wolf observes that one might question this claim "if one identifies Reason with certain concrete forms of thought and argument the relative value of which one questions." She also observes that ""Reason" is sometimes contrasted to emotion, for example, and associated with exclusive attention to precise logical argument and a preference for thinking in quantitative terms. A person who always consults and acts according to Reason in this sense might be found unattractively cold, straitlaced, lacking in spontaneity." However, Susan Wolf points out that these worries don't make sense if we take Reason to refer to "the highest faculty, or set of faculties, there are—that is, to whatever faculties are properly thought to be most likely to lead to true beliefs and good values." Thus, there is no way to offer rational criticism of an agent who always acts in accordance with his Reason. Or, as Wolf clarifies: "In light of that, any attempt to offer reasons for wanting to act against Reason will only show that the sense of Reason under attack is not the sense intended." Susan Wolf, *Freedom Within Reason* (Oxford: Oxford University Press, 1990), 56.

[133] Ibid. 57.

The problem is, however, that there is also reason to think that *having* the power to do otherwise when its exercise would be irrational is not something worth wanting. This is what Locke seems to suggest in this famous passage:

> Is it worth the Name of *Freedom* to be at liberty to play the Fool and draw Shame and Misery upon a Man's self? If to break lose from the conduct of Reason, and to want that restraint of Examination and Judgment, which keeps us from chusing or doing the worse, be Liberty, mad Men and Fools are the only Freemen: But yet, I think, no Body would chuse to be mad for the sake of such Liberty, but he that is mad already.[134]

Apparently, according to Locke, the ability to act irrationally constitutes a defect rather than a power worth wanting. For, madness and foolishness are certainly impairments of rational abilities. But if Locke is right about this, it is not clear how the ability to do otherwise (per se) could contribute to our control over our rational actions and ground our moral responsibility for them. The ability to do otherwise could ground responsibility for bad or irrational actions because the ability to act rationally is certainly not a defect. But, in that case, as proponents of the Reason View point out, what is relevant is not the ability to do otherwise per se but the ability to do the right thing for the right reasons.

So, consideration of the exercises of ability to do otherwise in particular cases gives us reason to be suspicious about the claim that moral responsibility requires the ability to do otherwise. Let us examine if this suspicion is well grounded.

4.2.2.1 Ability to Do Otherwise and the Value of Alternatives

[134] John Locke, *An Essay Concerning Human Understanding*, P.H. Nidditch, ed. Oxford: Clarendon Press, 1689/1975.

As I indicated at the beginning of the previous section, from the perspective of the Reason View, for the purpose of evaluating the ability to do otherwise, it is crucial to know what the quality of the particular alternative action is. To motivate that claim, I mentioned cases in which doing otherwise is irrational or opposed to Reason. However, one might object that that type of case cannot provide evidence for the claim that the ability to do otherwise is irrelevant when we act rationally. For an alternative to a rational action is not always an irrational action. This is so because we often have the *same* or *equally good* reasons for more than one action. This happens when we have to pick one out of many indistinguishable items in a supermarket, or one of several things that are indistinguishable as far as our interests are concerned (Van Inwagen calls the latter type of situation the 'vanilla/chocolate' cases). In fact, it seems that we have such alternatives all the time because we can always do things in different ways. For example, I can almost always use a different hand in pushing knob, or kick a ball with a different leg etc. In addition, we can almost always do things at a slightly different time. Finally, situations in which we just cannot decide which option is better are not so rare either. This is important because it is not clear that exercising these abilities is undesirable, or that having them is not valuable, or even that having them is some kind of defect.

However, in my view, this observation is not very significant because the alternatives which are not irrational cannot ground *moral* responsibility for rational actions. This is so because what we want to know when we inquire about an agent's responsibility is whether the agent was free to perform the *type* of action he actually performed (whether that was up to the agent); and the account of freedom in terms of alternative possibilities can answer that question only if it can point to an alternative type of action that the agent could perform, or the omission of the type of action he actually performed. Thus, since the type of action we are interested in is 'good action,' the alternative possibility which explains why the agent

performed that action freely must be either of the type 'not good' or 'bad' action (or omission). Otherwise, it is not clear how the theory in question could ground blameworthiness or praiseworthiness for an action. The following passage by Neil Levy testifies that I am not alone in thinking about moral responsibility in this way:

> for any action for which the agent is supposed to be directly free, the alternative action—the action he chooses in a significant proportion of nearby possible worlds—is an action with a conflicting moral valence (if the alternative action does not have a conflicting moral valence, then the action is not a locus of direct responsibility at all; it can at best be the locus of derived responsibility, where the derivation is from an action for which the agent is directly responsible). A moral valence, as I use the term here, is its polarity: an action has a positive valence if it is good, and a negative valence if it is bad; actions conflict in valence if one alternative is either good or bad, and the other is either of the opposite valence or morally neutral.[135]

One could perhaps argue that Levy and I are wrong in the following way. One could begin by noticing that people sometimes make choices between options that have the same moral value (e.g. when we have moral dilemmas), and that such choices are obviously moral choices. One could then notice that if they can do otherwise in such situations, their choices are free choices. Finally, one could ask: why aren't they morally responsible for making those choices and why we cannot say that the ability to do otherwise grounds their moral responsibility for those choices? My answer is that we cannot say that because our ability to do otherwise does not explain the fact that we have freely made a *moral* choice. That would be the case only if we could refrain from making a moral choice, or at least if we could make a less valuable choice. Thus, we must focus on the question whether the ability to something irrational (or bad) could ground our moral responsibility for our rational (or good) actions.

[135] Neil Levy, *Hard Luck: How Luck Undermines Free Will and Moral Responsibility* (Oxford: Oxford University Press, 2011), 42.

4.2.2.2 Ability to Act Irrationally and Ability to Act Crazily

When we consider alternative scenarios in which the agents don't act rationally, we see that in some of them agents act irrationally, but in some they act crazily. This is clear when we consider Luther cases. If I were to accept the offer to torture someone for a small amount of money (or to do that for any reason), that would not only be a sign of my irrationality, but rather of a serious impairment of my ability to appreciate reasons. And if the ability to do other than what the Reason suggest is the ability to act in such a way, it is really hard to see how having of that ability could ground moral responsibility for good actions. For, the ability to suffer a collapse of rationality is no ability at all. It is rather a liability.

It is plausible to assume that this is the sort of ability Locke had in mind when he said that only lunatics would want to have it. And, more generally, it seems that this picture of ability to act 'against the Reason' represents the main reason why some philosophers reject the relevance of ability to do otherwise as a condition of moral responsibility. This is certainly the case with Suzan Wolf who identifies agents who can act irrationally with autonomous agents described by her in the following way:

> They (autonomous agents) must be agents who not only *do* make choices on no basis when there is no basis on which to make them, but who also *can* make choices on no basis even when some basis is available. In other words, they must be agents for whom no basis for choice is necessitating. If the balance of reasons supports one alternative over all the others, it is still open to them to choose whether to act in accordance with the balance of reasons or not. We must now consider whether we have any reason to want to be autonomous agents.[136]

In this passage Wolf does not *identify* autonomy with the ability to act for no reason (when there is some reason), but she certainly suggests that being autonomous, among other things,

[136] Wolf, *Freedom Within Reason,* 55.

implies having that ability. And this seems to be enough to put to question the value of autonomy.

However, it is not clear that an agent who is able to act irrationally must be an autonomous agent in Wolf's sense. Experience seems to show that most people sometimes act irrationally, but in their 'normal' states they never act in a way that shows total insensitivity to rational considerations.[137] Thus, it seems plausible to say that most people can act irrationally but not autonomously. Wolf seems to identify the relevant ability with autonomy because she thinks that we value ability to act irrationally because of our desire to be free of all constraints even of those imposed on us by our own rationality.

This shows that it is not so easy to reduce the traditional view to absurdity. But, that does not eliminate the worry about the value of ability to do other than a good action. For, if we believe that agents in the Luther cases act with direct responsibility, the ability to do otherwise seems at least sometimes irrelevant to moral responsibility. In addition, there may be good reasons to think that ability to do otherwise is not necessary even when doing otherwise amounts simply to acting irrationally. I will argue that it is so bellow, but before I do that, I will consider a possible objection to the claim that ability to do something crazy cannot ground one's responsibility for doing something rational or normal.

4.2.2.3 Powers and Dispositions

The main reason why the ability to do something crazy (or to do something that is crazy in the given circumstances) seems inadequate for grounding moral responsibility is that we cannot see how its exercise can be exercise of control or even something that the agent

[137] It is not clear, though, that even the ability to act irrationally can survive if we pursue a logic according to which we evaluate abilities by considering why the agent performed this action rather than some other action with a different value. For a choice of even a slightly less rational course of action could perhaps be interpreted as a collapse of rationality.

does. But perhaps it is a mistake to evaluate abilities by imagining its exercise. Perhaps abilities do not have such a tight connection with counterfactual scenarios as we usually assume they have. This could be the case because there is no connection between abilities and certain kinds of counterfactual scenarios. Thus, it is impossible to imagine that a rational person acts crazily because crazy behavior is inconsistent with the person's rationality (conceived of as a character trait). But that may not be relevant for judging that he nevertheless has the ability to act in such a way. For the fact that it is unimaginable or inconceivable that the person will act crazily can be explained by pointing to the fact that the person lacks a *disposition* to act crazily (which is the reason why the person is rational), rather than by the person's lack of *power* to act in such a way. This explanation is possible because, as Ferenc Huoranszki points out, dispositions involve behavioral tendencies and imply possession of the relevant powers, but not the other way around.[138] One may have a power to do something without having any tendency to exercise that power. In that case it may be inconceivable that he would perform that action, but that does not mean that the agent lacks the power to perform that action, but only that he lacks a disposition to perform it.

So, the distinction between powers and dispositions may serve to explain how an agent can do something inconceivable. But can the same distinction be used to eliminate the worries about the value of having the ability to do otherwise in cases in which doing otherwise would signify the breakdown of the capacity to appreciate reasons? It may seem that it can, because if there is no possible scenario in which the agent exercises some ability, it is irrelevant how the exercise of that ability would look like. Thus, an ordinary person who is not a psychopath may have the power to accept the offer to torture an innocent person for a small amount of money, although there is no possible world in which the agent would exercise that ability in the normal circumstances. The exercise of that ability would indeed be

[138] See Ferenc Huoranszki, Powers, Dispositions and Counterfactual Conditionals, *Hungarian Philosophical Review*56 (2012): 33-53.

a sign of some defect in the agent; it would strongly suggest that the agent has gone crazy. But, the agent is not crazy merely because he is *able* to do something that would in the circumstances in question be crazy.

In my view, however, the distinction between powers and dispositions does not eliminate the worry about the relevance of ability to do something crazy (or something that is in the circumstances crazy) for moral responsibility. For, putting aside the question about the existence of abilities whose exercise is inconceivable, without being able to see what the exercise of some power would look like, we could not draw a distinction between *abilities* and *liabilities*.

Let me clarify this idea with an example. Consider the "ability" to make a mistake. Making a mistake is something that we are liable to doing; it can happen to us, but it is not correct to say that we have the power to make mistakes unless we are being ironic. Think about this simple algebra operation: adding 1 to 1. Imagine you have to choose between saying that the result is 2 and saying that the result is 3. You could say that it is 3 in the sense that you have everything you need to *say* 'three.' It seems that you have everything you need to *select* the wrong answer. But do you have the power to make a mistake? It seems not, because consciously choosing the wrong answer is not making a mistake. What we could say is that you might be liable to making mistakes. It could happen that something distracts you and you inadvertently select 3 instead of 2.

My point is that what the exercise of some capacity would be like really matters in determining whether we should categorize it as ability or as liability. To determine in which category some capacity belongs, we must put together in our minds the state of the agent before the action, keep it fixed, and add to it the occurrence of the action in question.

However, one might object that my example is misleading. Making a mistake in algebra is not a matter of choice, because succeeding in it is not a matter of choice either.

When we work on mathematical problems, we are trying to find a solution and at some point the solution just appears before our minds eye. Or if we have made some mistake in the process, the wrong solution appears before our minds eye. It is never the case that we have to *decide* whether the solution is right after we perceive it. We either believe that it is right or not. The same is the case with other sorts of mistakes. When one trips, one is usually not asked beforehand to decide if he or she will trip or not trip. However, refusing to do something crazy can be a matter of choice. Thus, the line between abilities and liabilities could perhaps be drawn by using the distinction between things that can be and those that cannot be a matter of choice.

But, we may also conclude from this that the line between some abilities and liabilities cannot be drawn by considering whether their manifestation can be a matter of choice. Some failures to exercise our abilities can be a matter of choice (although they are not a matter of choice under that description). In ordinary cases we cannot choose to make or not to make a mistake, that is, we cannot choose whether to succeed or fail in the exercise of our abilities because most ordinary abilities are abilities that we can perform intentionally. And their success is measured by whether they were performed intentionally. But the same criteria cannot be used for evaluating capacities that cannot be exercised intentionally which is the case with making a particular choice.[139]

Thus, it seems that the only sort of ability to do otherwise that may ground our responsibility is the ability to act irrationally or less rationally, i.e. akratically. Let us see whether we have good reasons to want this ability.

[139] It is difficult to see how making a particular choice can be an intentional action if to be intentional action must be preceded by an intention to perform that action. Hugh McCann, however, argues that choices don't require prior intentions because they are intrinsically intentional actions. See: Hugh McCann, *The Works of Agency: On Human Action, Will, and Freedom* (New York: Cornell University Press, 1998), 92.

4.2.2.4 Ability to Do Otherwise and Self-Determination

One of the main reasons why philosophers think that moral responsibility requires ability to do otherwise is that it requires self-determination. Plausibly, to be truly responsible for what we do it is not just enough that we are able to act voluntarily or on the basis of our choices; to be truly responsible we must also be able to determine the contents of our wills. And it is natural to assume that for that purpose we must have the ability to do or to will otherwise. But is the ability to do otherwise really necessary for self-determination? And, can't we say for an agent whose actions are determined by his Reason that he has power of self-determination?

An interesting argument for the claim that self-determination requires ability to do otherwise has been presented by Thomas Pink. According to Pink, the accounts according to which some sort of rationality or reasons-responsiveness is sufficient for moral responsibility cannot explain how self-determination is possible because rationality or responsiveness to reasons is not a *power* to determine for oneself what to do, but rather a *mode* of determining of what we do. In fact, in his view, rationality is not the ability to *determine*, but to *be determined* by what reasons there are. Rationality applies to our relation to our beliefs (or relation between our beliefs and justifications of those beliefs). If we are rational, our beliefs will be determined by the evidence that we have for them and that excludes our control over them. Thus, according to Pink,

> the function of rationality in relation to such belief is then to ensure that the capacity to be determined which such belief involves functions properly – in response to what the justifications presented to me really are. Rationality ensures that, far from what I believe being left for me to determine, my beliefs

> faithfully track experience and the evidence it provides, and are determined by it.[140]

If we describe rationality in this way, Pink is obviously right. Rationality in this sense has *nothing* to do with self-determination. It is at best a description of the way we use the power of self-determination. But, it is not obvious that rationality has nothing to do with self-determination if we understand self-determination (or one aspect of it) as the ability to *act* in accordance with reasons (rather than a capacity to be influenced by reasons).[141] In particular, that is not clear if we assume that the exercise of that ability is determined by our Reason rather than by some blind internal or external forces. For as Suzan Wolf notices,

> The position we are considering assumes that one's freedom of choice would be compromised if one's choice necessarily followed one's Reason. It assumes that insofar as one's Reason is unconditionally decisive in determining one's choice, to that extent the choice is not truly and ultimately one's own. *These assumptions reveal an implicit conception of Reason as alien to oneself, as a determining force with which one might in principle be in competition* (my italics). But, holding fast to the broad and essentially normative use of the word Reason, it is not clear that such a view is intelligible.[142]

In my view, this is a very powerful reply. However, it fully eliminates the worry about the lack of self-determination only if we identify self-determination with the determination of our will by our Reason. In that case, the question remains about self-determination in cases when our Reason fails to determine our actions. To account for self-determination in those cases ability to act for good reasons does not seem sufficient. For self-determination in those cases seems to require the ability to exercise or not exercise the ability to act for good reasons. But, if we think that self-determination consists in our acts of allowing or not allowing our will to be determined by our Reason, then the Reason view has a problem of explaining how self-

[140]Thomas Pink, "Power and Moral Responsibility," *Philosophical Explorations*12:2 (2009): 138.

[141] One might object that there is no such ability as the ability to *act* for reasons, but rather just capacities to do certain things (e.g. raise a hand, sing, laugh etc.) that may or may not be exercised in accordance with reasons.

[142] Wolf, *Freedom Within Reason*, 58.

determination is possible. For such self-determination requires a two-way power distinct from the one-way power to act in accordance and on the basis of good (or right) reasons.

In my view this consideration shows that the proponent of the Reason View has a good reason to say that it is impossible to act freely contrary to Reason. That is, they have reason to identify free action with action that is determined or flows from one's Reason. This way they could eliminate the problem with self-determination without giving up their main claim that free will is essentially ability to act in accordance and on the basis of good reasons. For, it seems plausible to say that agents who act for good reasons have the ability to act for good reasons.[143]

The plausibility of this version of the Reason view will be the topic of the next chapter. Here I just want to consider if there is some other reason why we should think that in addition to the ability postulated by the proponents of the Reason View self-determination or moral responsibility requires ability to do otherwise.

4.2.2.5 Free versus Automatic Action

Earlier in this chapter, in the section on what I call Luther cases, I mentioned a story by Susan Wolf about a woman who saved a drowning child. According to Wolf, the woman is praiseworthy for her action even though she could not do otherwise. In responding to the objection that this woman cannot be morally responsible unless she could have done otherwise, Wolf introduces into the story another woman who is the same as the first woman in all relevant respects except that she, unlike the first woman, could do otherwise. She asks then whether we should consider only the second woman morally responsible for saving the child, and if so, why.

[143] This no doubt sounds like a version of a Real-Self View, in which the real self is the agent's Reason.

Wolf observes that one might think that the first woman could not be morally responsible because her lack of ability to do otherwise indicates (as I suggested earlier) that she acted *automatically* or from an obsessive or blind impulse. In other words, she notices that one might think that the lack of ability to do otherwise excludes possibility of acting on the basis of reasons. However, according to Wolf, this suggestion fails because "mechanical action is properly opposed not to autonomous but to rational action, and the women in question differ in autonomy but not in rationality".144

Wolf acknowledges, however, that her example might not be a perfect illustration of this point because it describes an emergency situation in which a "near-reflex action is needed"[145]. As she observes, that may be the reason why the example does not reveal that the ability to do otherwise is necessary for exercising "subtle powers of discrimination and refined faculties of judgment?"[146] But, as Wolf points out, this characteristic of the example is irrelevant. For if

> we move to a non-emergency example in which time for reflection and deliberation is available, we can see that again the non-autonomous agent need act no more mechanically than the autonomous agent; the actions of the former may be as finely cued to subtle perceptions and sophisticated patterns of reasoning as those of the latter. For, again, the difference between the autonomous and nonautonomous agents lies not in their capacities to use Reason, but in their capacities to reject Reason.[147]

As I mentioned earlier, I find Wolf's identification of ability to do other than the (or a) right action for the right reasons with autonomy or ability to reject Reason very problematic. Nevertheless, I think that this passage contains a good point. There is no reason to think that the person whose actions are determined by her Reason cannot act for reasons, because it is

[144] Wolf, *Freedom Within Reason*, 60.
[145] Ibid.
[146] Ibid.
[147] Ibid., 61.

not clear how merely not being determined by Reason could provide the power to act for reasons to someone who otherwise lacks that power.[148]

However, the problem might not be that the agent who is deprived of ability to do otherwise could not act for a reason but that he could not do it freely. For, even the compulsive agents can act for reasons, but they don't act freely and cannot be morally responsible, and the explanation for that seems to be that they cannot do otherwise.[149]

But, it is difficult to see how the notion of compulsion can be applied to the case of doing the right thing for the right reasons. The idea is intelligible only if we could get into collision with our own Reason, if we could become prisoners, so to say, to our own rational capacity. But that seems impossible, because as soon as we recognize our behavior as unfree due to our 'Reason,' the inclination that comes from that source would not anymore be a product of our Reason. And it seems to be a feature of compulsive behavior that compulsive agents themselves find it problematic.[150] The collision seems possible though in akratic cases, but I think that we should not take it for granted that these cases are possible.

4.3 Conclusion

In conclusion, I think that we have very good reasons to accept Suzan Wolf's Reason View and her claim that free will is asymmetric in the sense that it involves ability to do otherwise when doing bad things is concerned but not when doing bad things is concerned.

[148] Someone could perhaps question the assumption that Reason could determine one's actions. But Luther cases seem to support that assumption. More precisely, they seem to support the assumption that Reason can determine one's behavior assuming also that nothing unusual will happen to the agent, that the agent will not suffer a nervous breakdown or suddenly feel an overpowering desire to do otherwise. If there is no chance that the agent will do otherwise in these circumstances why can't we say that Reason together with other factors determines the agent's behavior?

[149] I am grateful to Ferenc Huoranszki for drawing my attention to this objection.

[150] Perhaps we can say that akrasia is an example of the collision between a person and her Reason. But the experience of what seems to be akratic behavior does not show that. It shows rather that the agent acts in spite of herself or at best that the agent is in conflict with oneself.

The first part of this asymmetry thesis follows from the comparison of Wolf's Reason View with the other 'rationalist' views which imply that moral responsibility never requires ability to do otherwise. For, the comparison shows that the Reason View offers a more natural account of the rational capacities relevant for moral responsibilities than the rival views, and a more comprehensive explanation of our intuitions about moral responsibility in particular cases. In addition, unlike the other rationalist views, the Reason View seems to cohere better with some very plausible abstract ethical principles such as the principle that it is not fair to blame someone who could not do otherwise or the 'ought implies can' principle.

On the other hand, examination of the value of ability to do otherwise in particular cases vindicates the second part of Wolf's asymmetry thesis - the claim that moral responsibility for good actions does not require ability to do otherwise. For, the kind of ability to do otherwise that seems relevant for moral responsibility could be either the ability to do something crazy or the ability to do something foolish/less worthy and it is not clear how these abilities could ground moral responsibility. The former ability cannot do that because it is not ability at all but rather liability. The latter ability does not seem relevant because it is difficult to see why an agent who has it would have more control over his behavior, than someone who does not have it. This is difficult to see because there seem to be no reason to think that the action of the former agent would be more rational, less automatic or more attributable to him than the of the latter agent, and there seems to be no question begging argument for the claim that the latter agent could not be rational or able to act voluntarily in the first place. The only reason to think that something more than the ability to do the right thing for the right reason is required for free will and moral responsibility is the requirement of self-determination. For, an agent seems to satisfy this requirement only if it is up to him whether he uses his power to act for the right reasons (when he recognizes them) and for that to be the case it seems that he needs the ability to do otherwise. However, this

problem can be avoided, as I have suggested, essentially by identifying the agent with his Reason and identifying free and responsible behavior with behavior determined by one's Reason, and by denying that there is such a thing as unexercised ability to recognize and act for the right reasons. For, in that case it is literally true that the agent's action is self-determined.

CHAPTER 5: TWO ARGUMENTS FOR THE VIEW THAT FREE AND RESPONSIBLE AGENTS CAN DO ONLY RIGHT THINGS FOR THE RIGHT REASONS

At the end of the previous chapter I suggested that the proponent of the Reason View should accept the claim that our actions are free only when they are determined by our Reason. I suggested this because that seemed to be the only way for the proponent of this theory to provide a satisfactory account of self-determination. In this chapter I continue to develop this suggestion by arguing that without its acceptance, the proponent of the Reason View cannot eliminate the worry of the so called source-incompatibilist that *determinism* undermines moral responsibility by precluding one from being the appropriate source of one's action. More precisely, I argue that without accepting my suggestion, the proponent of the Reason View cannot satisfactorily answer the source-incompatibilists challenge based on the similarity of certain sorts of manipulation that intuitively undermine moral responsibility with ordinary causal origins of actions in deterministic worlds.

But before discussing this source-incompatibilist challenge, I argue that my conclusions in previous chapters together imply that we cannot be morally responsible when we fail to do right things for the right reasons (for the sake of simplicity in the rest of this chapter I will call the actions that do not fall into this category 'bad actions.') . More precisely, I will show that the view in question follows from the combination of incompatibilism about the ability to do otherwise and determinism, which I defended in the first chapter, Asymmetry, and my conclusion of the third chapter that indeterminism is not relevant to free will, i.e. that libertarian free will has no more value than compatibilist free will.

5.1 An Argument for Skepticism about Responsibility for Wrong Actions

To see how this conclusion follows from my earlier conclusions, consider the view that R. Jay. Wallace calls 'selective incompatibilism.'[151] According to this view, determinism is incompatible with moral responsibility for bad actions but not with responsibility for good actions. It is the result of combining asymmetry of the Reason View and incompatibility of ability to do otherwise and determinism.

This view is no doubt a bit surprising. Susan Wolf, for instance, observes that if the Reason View had this implication it would be "very hard to swallow, and might well make one wonder whether our ordinary notion of responsibility were still being discussed."[152] However, to the best of my knowledge, no one has so far argued that this view is incoherent. Wallace rejects it because he rejects Asymmetry while Dana Nelkin and Susan Wolf reject it because they think that incompatibilism is false. But neither of them thinks that this view could not be true.

However, I think that this position *is* incoherent unless it is true that we can be responsible *only* for good actions performed for good reasons. For, consider what would be the case if it were otherwise. In that case, it would be possible that a person does something bad and deserve blame for what she has done in a world with indeterministic laws of nature are, but that in a world with deterministic laws the identical person does not deserve blame for the identical action. If we accept, as I did in the third chapter, the claim that mere indeterminism cannot increase one's free will and one's abilities in general, we must conclude that this scenario is simply unintelligible. For, mere indeterminism would have to

[151] I am not sure whether Wallace uses exactly this label anywhere, but he talks about 'selective incompatibility.' See R. Jay. Wallace, *Responsibility and Moral Sentiments*, 203.
[152] Susan Wolf, *Freedom within Reason*, 97.

account for the difference between the agent who has the abilities that the Reason View requires and the agent who lacks those abilities, and it is not clear how indeterminism could do that. The scenario in question would be intelligible only if the relevant rational abilities required some further power logically incompatible with determinism (e.g. agent-causal power) or if these abilities themselves were logically incompatible with determinism.[153] But neither of these two claims sounds credible. For it seems absurd to say that the truth of determinism implies that no one ever recognized the right reasons or acted on the basis of them, which would have to be the case if one of those claims above were correct.[154] Besides, it is not clear whether the existence of agent-causal power can be combined with the asymmetry of the Reason View. Therefore, we must conclude that the scenario mentioned above is impossible. And, in the light of what I have said so far, the best way to explain its impossibility is to reject the possibility of moral responsibility for bad actions, or possibility of justified blame.

However, one might object that instead of being a reason for skepticism about responsibility for bad actions, the unintelligibility of this sort of scenario in fact constitutes a reason against incompatibilism, because *incompatibilism* alone entails the possibility of such scenarios. For, one might argue that this is so because incompatibilists are committed not just to the claim that free will does not exist in any possible deterministic world, but also to the claim that determinism makes a difference to whether free will exists, i.e. that in deterministic worlds agents lack free will *because* of determinism. One could support this last claim by pointing out that if that was not so, it would be impossible to distinguish incompatibilists from impossibilists (philosophers who think that free will is impossible). For the latter agree with incompatibilists that there is no free will in deterministic worlds, but for

[153] For, in that case, it would be true that without indeterminism one could not have some power that constitutes one's free will.
[154] The idea that agency requires indeterminism is also relevant here.

reasons that may have nothing to do with determinism. Thus, perhaps the idea of difference-making that seems essential to incompatibilism requires the possibility scenario of the above mentioned sort.[155]

This objection to my argument, however, rests on a mistaken conception of incompatibilism. Incompatibilism is simply the thesis that there is no possible world in which determinism is true and in which someone has free will (or ability to do otherwise in this case).This thesis says nothing about the reasons why people in deterministic worlds lack free will in particular cases. In particular, it does not say (nor does it imply) that deterministic natural laws (themselves) deprive agents of free will in those worlds, although their lack of free will has something to do with the deterministic nature of those laws. If it were otherwise, it would not be possible to count someone who believes in agent-causation as incompatibilist because it makes no sense to talk (except metaphorically) about determinism depriving someone of his agent-causal power. Rather the connection between agent-causal power and determinism is logical. The same is true if we talk about the relation between determinism and ability to do otherwise. The incompatibility between them is a matter of logical relations revealed by the Consequence Argument. Consequently, the incompatibilist is not someone who believes that deterministic laws *deprive* agents of free will, but simply someone who thinks that there is no space for free will in deterministic worlds because of the logical relations between determinism and free will.

Another objection that one might raise against my argument is that the view I defended in the previous chapter is not in fact asymmetric. That is, one might argue that I myself reject one of the premises of my argument (other than the one that I want to reduce to

[155] I am not sure that anyone would actually argue against my argument in this way, but some passage in texts by Kadri Vihvelin and Kristin Mickelson (Demetriou) indicate that they may be inclined to do that. See Kadri Vihvelin *Causes, Laws, and Free Will: Why Determinism Doesn't Matter* (Oxford: Oxford University Press, 2013), 23-35. Kristin Mickelson, "A Critique of Vihvelin's Three-Fold Classification," *Canadian Journal of Philosophy* 45 (2015): 85-99

absurdity). One might argue for that claim by pointing out that no asymmetry follows from the view that moral responsibility essentially depends on determination of our actions by our Reason.[156]

This objection is interesting, but it also fails because it is not true that asymmetry does not follow from the above mentioned claim. For it is an essential element of the determination in question that the agent has the relevant rational abilities.[157] Consequently, to be blameworthy one would also have to possess those abilities and for that one would have to be able to do otherwise. One could criticize this reasoning by saying that something that is impossible cannot require something else for its existence. But that does not seem correct. The fact that a square circles are impossible does not imply that it is not a requirement for a circle to be square that it has four angles. In fact, exactly because such a requirement exists we know that square circles are impossible.

Finally, one might say that my argument is circular. For, I argued in the previous chapter that in order to defend the Reason View from the objection that it does not provide an account of self-determination, we must assume that the unexercised ability to do the right thing for the right reason is not possible. This is true. But, I don't think this is a problem for my argument, because the argument is meant to be simply an alternative way of persuading those who *already* accept Reason View that it has the implication that I argue it has.

Therefore, since I don't see any other objection that one could raise against my argument, I conclude that what I said previously supports the claim that no one can do bad things freely and be morally responsible for doing bad things. But, what should we say about responsibility for good actions? Can we conclude on the basis of my conclusions in the previous chapters that such responsibility is possible and that it in fact exists? Clearly, my

[156] I am grateful to my supervisor Ferenc Huoranszki for drawing my attention to this problem.
[157] It is important to notice also that the determination relation in question is not something over and above the possession of rational abilities in question. It is just a consequence of the fact that these abilities cannot be unexercised.

conclusions imply that determinism cannot create problems for moral responsibility for good actions by eliminating alternative possibilities because I concluded that moral responsibility for good actions does not require alternative possibilities. In addition, as I said above, it does not seem that determinism per se implies that no one has the ability to recognize and act for good reasons. Furthermore, as we have seen earlier, indeterminism also does not represent a threat to moral responsibility especially when what we do is supported by good reasons. However, there is another concern about determinism which seems independent of the concerns so far mentioned. It is the concern that if our actions follow deterministically from what happened in the past, our actions cannot originate from us in the way necessary for moral responsibility. In the rest of this chapter I will address this worry.

5.2 The Manipulation Argument(s)

As we have seen in the previous chapter, the rationalist views of free will which do not require ability to do otherwise for moral responsibility seem to be under pressure to provide some account of the origin of the processes or structures that lead responsible agents to their actions. In other words, it seems that these philosophers must acknowledge the relevance of history for moral responsibility; they must hold that moral responsibility is a historical phenomenon. Otherwise, they must accept the consequence that even agents who are manipulated in certain ways (e.g. the victims of brainwashing) can be morally responsible for their actions. For, there may be no non-historical difference between the agents that satisfy their conditions for morally responsible action and agents who are not morally responsible because they are victims of manipulation. This is so because there is no reason to

think that the structures or processes which, according to these theories, lead morally responsible agents to their actions cannot be products of manipulation by other agents.[158]

For this reason many philosophers recognize the importance of causal histories of actions to the question of moral responsibility. But, the question is whether it is possible to acknowledge this without having to accept incompatibilism. For, it may be that there is no relevant historical difference between the actions of manipulated agents and of any other agents in deterministic worlds. It might be that the only thing that is relevant for judgments of moral responsibility of both types of agents is that their actions are determined by factors over which they have no control. This is precisely what the arguments called the 'manipulation arguments' are designed to show. In what follows, I will present two most famous arguments of that sort: Derk Pereboom's Four-Case Argument and Alfred Mele's Zygote Argument.

5.2.1 The Four-Case Argument

A common feature of all manipulation arguments is that they are based on certain cases involving agents manipulated in certain ways. These cases are supposed to support the main idea of the arguments that there is no relevant difference between causal determinism and responsibility-undermining sorts of manipulation. The distinctive feature of Pereboom's manipulation argument is that it is based on four cases which include three manipulation cases and one ordinary deterministic case. Each of the cases in question involves an agent, Mr. Plum, who is causally determined to kill and eventually kills another agent, Ms. White. In addition, in each of the cases, Mr. Plum is such that he satisfies the (non-historical)

[158] Some philosophers accept this consequence. Thus, Harry Frankfurt says the following: "It is possible that a person should be morally responsible for what he does of his own free will and that some other person should also be morally responsible for his having done it." Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68 (1971): 20. However, this view is general considered 'bullet-biting.'

conditions for moral responsibility postulated by Frankfurt, Hume and Ayer, Fischer and Ravizza, and R.J Wallace. More precisely, he has the relevant combination of first order and higher order desires ("his desire to kill White conforms to his second-order desires in the sense that he wills to kill and wants to will to kill, and he wills to kill because he wants to will to kill."[159]). In addition, his desire to kill Ms White is not irresistible; the mechanism which leads to his action is reason-responsive ("if he knew that the harmful consequences for himself resulting from his crime would be much more severe than they are actually likely to be, he would not have murdered White."[160]) and his action is caused "by desires that flow from his "durable and constant" character."[161] Finally, he has the general capacity to grasp the relevant reasons and regulate behavior by their light (whenthe egoistic reasons that count against acting morally are relatively weak,he will typically regulate his behavior by moral reasons instead."). Pereboom's aim is to show that although Plum has all these features, he is not morally responsible for killing Ms White in the scenario in which his action has an ordinary causal history (in Case4) because there is no relevant difference between ordinary causal determination and manipulation.

Pereboom presents the cases in question in the following way:

> Case1. Professor Plum was created by neuroscientists, who can manipulate him directly through the use of radio-like technology, but he is as much like an ordinary human being as is possible, given this history. Suppose these neuroscientists "locally" manipulate him to undertake the process of reasoning by which his desires are brought about and modified – directly producing his every state from moment to moment. 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. Plum is not constrained to act in the sense that he does not act because of an irresistible desire- the neuroscientists do not provide him with an irresistible desire- and he does not think and act contrary to character since he is often manipulated to be rationally egoistic. His effective first-order desire to kill Ms. White conforms to his second-order desires. Plum's reasoning

[159] Derk Pereboom, *Living Without Free Will*, 111.
[160] Ibid.
[161] Ibid

process exemplifies the various components of moderate reasons responsiveness. He is receptive to the relevant pattern of reasons, and his reasoning process would have resulted in different choices in some situations in which the egoistic reasons were otherwise. At the same time, he is not exclusively rationally egoistic since he will typically regulate his behavior by moral reasons when the egoistic reasons are relatively weak- weaker then they are in the current situation.[162]

Case2. Plum is like an ordinary human being, except that he was created by neuroscientists, who, although they cannot control him directly, have programmed him to weigh reasons for action so that he is often but not exclusively rationally egoistic with the result that in the circumstances in which he now finds himself, he is causally determined to undertake the moderately reasons responsive process and to possess the set of first- and second-order desires that results in his killing Ms. White. He has the general ability to regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and accordingly he is causally determined to kill for these reasons. Nevertheless, he does not act because of an irresistible desire.[163]

Case3. Plum is an ordinary human being, except that he was determined by the rigorous training practices of his home and community so that he is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1 and 2). His training took place at too early an age for him to have had the ability to prevent or alter the practices that determined his character. In his current circumstances, Plum is thereby caused to undertake the moderately reasons-responsive process and to possess the first- and second-order desires that result in his killing Ms. White. He has the general ability to grasp, apply and regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and hence, the rigorous training practices of his upbringing deterministically result in his act of murder. Nevertheless, he does not act because of an irresistible desire.[164]

Case4. Physicalist determinism is true, and Plum is an ordinary human being, generated and raised under normal circumstances, who is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1-3). Plum's killing of White comes about as a result of his undertaking the moderately reasons-responsive process of deliberation, he exhibits the specified organization of first- and second-order desires, and he does not act because of an irresistible desire. He has the general ability to grasp, apply, and regulate his behavior by moral reasons, but in these circumstances the egoistic reasons are very powerful, and together with background circumstances they deterministically result in his act of murder.[165]

Pereboom starts his argument with the observation that it is obvious that Plum is not morally responsible for killing Ms. White in Case 1. This is so, in his view, because Plum was

[162]Derk Pereboom, *Living without Free Will*, 112-113
[163] Ibid. 113-114.
[164] Ibid.
[165] Ibid., 115.

causally determined to kill Ms White "by the neuroscientist's activities which are beyond his control."[166] Nevertheless, Pereboom admits that some readers may have a feeling that Plum's non-responsibility is due to the way in which he was manipulated (i.e. direct or local manipulation) rather than to causal determination.

According to Pereboom, Case 2 is supposed to eliminate this feeling. For, in his view, this case shows that neuroscientists can accomplish the same effect without direct stimulation of Plum's brain, that is, without "producing his every state from moment to moment."In addition, comparison with Case1 shows that manipulation does not have to be direct to undermine one's responsibility. For, as Pereboom observes, "whether the programming takes place two seconds or thirty years before the action seems irrelevant to the question of moral responsibility"[167] From that he infers that just like in Case 1 we should not hold Plum morally responsible for killing Ms White in Case2.

Now, as Pereboom points out, the only difference between Case3 and Case2 is that in the former case Plum's act of murder was not causally determined by intentional activities of other beings. In other words, the only difference between these cases is that Case3 does not involve manipulators. However, according to Pereboom, this difference is not relevant for judgments of responsibility because replacing manipulators with blind forces or "randomly created machines" would not eliminate the intuition that Mr. Plum is not responsible.[168] Consequently, he concludes that Mr. Plum is not morally responsible for killing Ms White in Case3 either and that the best explanation why that is so is causal determination of his action by factors over which he had no control.

Finally, since there is no difference between cases 3 and 4, (except that it is clear what causally determined Mr. Plum to kill Ms White in Case3), and Case4 is just an ordinary

[166] Ibid.,113.
[167] Ibid. 114.
[168] Ibid. 115-116.

deterministic scenario, Pereboom concludes that Plum is not morally responsible for killing Ms White in this case either and that the best explanation of why it is so is causal determination of his action by factors over which he did not have control.

Pereboom's argument can be summarized in the following way:

1) Mr. Plum is not morally responsible for killing Ms White in Case 1 because of the way he was manipulated.
2) There is no relevant difference between the histories of Mr. Plum's actions in cases 1 - 4 as far as his moral responsibility is concerned.
3) The best explanation of why Mr. Plum is not morally responsible for killing Ms White in cases 1-3 is that his action was causally determined by factors over which he did not have control.

Therefore,

4) Plum is not morally responsible for killing Ms White in Case4 and the fact that his action is determined by factors over which he had no control is the best explanation of why he is not morally responsible.

This argument is obviously valid, but one may argue that it is not sound. Most critics have so far concentrated on the first and second premises of the argument. The attack on the first premise is known as the hard-line reply, while the attack on the third is known as the soft-line reply. These labels are due to the fact that it is intuitively more difficult to accept the claim that manipulation and moral responsibility are compatible than the claim that there is a significant difference between manipulation and causal determination.

I believe that a convincing soft-line reply to Pereboom's argumentis available, but in my view, hard-line reply is what is ultimately required to eliminate the worries about

manipulation. I believe that a special advantage of the Reason View, and in particular my version of this view over other source compatibilist views is that it has resources for giving such a reply.

In the next two sections I examine and reject the soft-line strategy for dealing with manipulation arguments.

5.2.2 Soft-Line Objection to the Four-Case Argument

Virtually everyone is inclined to give a soft-line reply to Pereboom's argument in so far as Case1is concerned. In other words, most philosophers think that there is a significant difference between the way in which Plum is manipulated in this case and ordinary causal determination. This is so, as Michael McKenna points out, because Plum in this case looks more like a cartoon character drawn from moment to moment than a genuine human being. Similarly, Fischer and Ravizza notice that it is difficult to see how Plum can be "a coherent self" or "a genuine self", because "from the beginning, there has been no opportunity for a genuine self to emerge and develop."[169]

What about Case 2? There is no reason to worry here that Plum is a person and that *he* acts since the manipulators do not influence his decision making process directly. Thus, assuming that he really satisfies the non-historical compatibilist conditions, the non-historical compatibilist must admit that Plum is morally responsible in this case. In other words, it seems that the non-historicist compatibilist must adopt a hard-line approach to this case. However, the historicist compatibilist can plausibly argue that Plum is not morally responsible because he does not have an appropriate history. For, determining one's attitudes from a (considerable) temporal distance seems possible only if it involves 'bypassing' of the

[169] Fischer and Ravizza, *Responsibility and Control*, 230-235.

agent's capacity for critical examination of his own attitudes and such bypassing is not a necessary effect of ordinary causal determination. Alternatively, if the Plum's history is such that his capacity for critical reflection was not bypassed, the historicist compatibilist could say that his act was not in fact a result of manipulation and that he is responsible for his act of murder.[170]

Similar reply is available concerning the Case 3.For, it is implausible to assume that the normal influences of upbringing and of social environment could be so strong to determine an agent's actions in distant future by determining his attitudes. Education and upbringing seem to have this effect only in pathological cases.

Therefore, soft-line replies seem to be available for each of the cases of manipulation presented by Pereboom. In other words, it seems that Pereboom has not shown that there is no difference between (the responsibility-undermining) manipulation and causal determinism.

However, according to Michael McKenna, Pereboom's cases can be modified to satisfy the historicist-compatibilist conditions on free agency.[171] As McKenna points out, this should be possible because psychological antecedents of all actions in deterministic worlds causally originate in factors external to the agent. In his view, the problem with Pereboom's cases is that they are under-described. So, according to McKenna, soft-line objections to Pereboom's cases can be eliminated by presenting more detailed versions of his cases. McKenna makes an attempt to show that. But instead of analyzing his versions of

[170] This may be a problem for what I say bellow about the Zygote argument.

[171] See Michael McKenna, "A Hard-line Reply to Pereboom's Four-case Argument *Philosophy and Phenomenological Research*77 (2008).

Pereboom's cases, I turn to Alfred Mele's Zygote Argument, which, in my view, shows even more clearly the inadequacy of soft-line replies to manipulation arguments.[172]

5.2.3 The Zygote Argument

In order to show that there is no relevant difference between an ordinary causally determined action and action causally determined by intentional activity of other agents, Alfred Mele presents the following story:

> (Goddess) Diana creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. From her knowledge of the state of the universe just prior to her creating Z and the laws of nature of her deterministic universe, she deduces that a zygote with precisely Z's constitution located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E. If this agent, Ernie, has any unsheddable values at the time, they play no role in motivating his A-ing. Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator. So he satisfies a version of my proposed compatibilist sufficient conditions for having freely A-ed.
> Compare Ernie with Bernie, who also satisfies my compatibilist sufficient conditions for free action. The zygote that developed into Bernie came to be in the normal way…[173]

According to Mele, this story supports the following argument:

1. Because of the way his zygote was produced in his deterministic universe, Ernie is not a free agent and is not morally responsible for anything.

[172] The problem is, however, that it is more accurate to describe Mele's argument as an 'original design argument,' than as a manipulation argument. The reason why he thinks so is apparently the fact that in the example that supports this argument there is no 'bypassing' of the agent's capacities for self-control.
[173] Alfred R. Mele, "Manipulation, Compatibilism, and Moral Responsibility," *Journal of Ethics* 12 (2008): 279.

2. Concerning free action and moral responsibility of the beings into whom the zygotes develop, there is no significant difference between the way Ernie's zygote comes to exist and the way any normal human zygote comes to exist in a deterministic universe.
3. So determinism precludes free action and moral responsibility.[174][175]

Premise 2 of this argument seems clearly true. For, as Mele points out, "a proponent of the Zygote Argument might contend that, given the additional facts that, in both universes, Ernie has no say about what causes Z, no say about the rest of the universe at that time, and no say about what the laws of nature are, the cross-universe difference in what caused Z does not support any cross-universe difference in freedom or moral responsibility."[176]

Mele's argument is thus immune to soft-line responses. Consequently, the crucial question is whether hard-line response to his argument could succeed. That is, it is crucial to see whether the fact that Ernie's action was determined by Diana entails that he was not morally responsible for that action and whether intentional determination of actions by other agents in general entails lack of moral responsibility (or free will relevant for moral responsibility).

According to Mele, the answer that one is likely to give to this question depends on one's view of free will before being presented with the argument. If one is already an incompatibilist, one will have the incompatibilist intuition, (i.e. one will find the premise true), and if one is a compatibilist one will intuit that this premise is false. According to Mele,

[174] Ibid. 280.
[175] In fact, as Kristin Mickelson points out, this argument is not valid. For all that follows from the premises of this argument is the conclusion that there are no free and morally responsible actions in deterministic worlds. Mele's conclusion follows only if we add the "diagnostic premise" that causal determinism represents the best explanation of Ernie's lack of moral responsibility. See Kristin Mickelson, "The Zygote Argument is Invalid: Now What?"*Philosophical Studies*172 (2015): 2911-2929.
[176] Mele, "Manipulation, Compatibilism, and Moral Responsibility, 280.

an unbiased and thus more valuable response is expected from an agnostic (someone who is undecided between compatibilism and incompatibilism). Mele reports that he is one and that he feels pull toward the truth of the premise 1.[177]

However, there is a strategy that puts to question premise 1of Mele's argument which does not rest only on our intuitions about particular cases. The strategy in question has been introduced by Michael McKenna as a hard-line reply to Pereboom's Four-Case Argument. But, as we shall see, it can be used as a reply to any manipulation or original design argument.

5.2.4 McKenna's Hard-line reply to the Four-case Argument

McKenna's strategy rests on his insight that the first premise of Pereboom's argument can be put to question by 'running the argument backwards.' To show this, McKenna starts by pointing out that the only warranted attitude concerning the responsibility of agents in ordinary deterministic scenarios (such as Pereboom's Case 4) is the agnostic attitude. The content of this attitude is that it is not clear whether the agent is morally responsible for his action. He then shows that if we accept the claim that there is no relevant difference between causal determination and manipulation, we must ultimately conclude that it is not clear whether the agent is responsible (or not responsible) for actions resulting from manipulation. In other words, the same reasoning (presented by Pereboom and Mele) which leads to the worry that we are not morally responsible if determinism is true if we start from the assumption that manipulation undermines moral responsibility, leads to agnosticism about incompatibility of manipulation and moral responsibility if we start from the agnostic attitude

[177] Mele adds, though, that he might not be a truly adequate agnostic since his agnosticism depends to a large extent on his optimism about the prospects of indeterministic free agency. Ibid. 280-283.

about responsibility in ordinary deterministic circumstances. Consequently, we cannot accept the first premise of Pereboom's argument and the argument cannot get off the ground.[178]

However, as Pereboom observes, the plausibility of McKenna's criticism depends on the specific understanding of the agnostic attitude that is rational to have concerning the agents determined in ordinary ways. According to Pereboom, his criticism is correct if the rational attitude is the attitude of what he calls the "confirmed agnostic". This type of agnostic is undecided about moral responsibility of agents in ordinary causally deterministic scenarios, but in addition to that believes that the issue is closed, and that no further considerations should sway him to one side or the other. However, according to Pereboom, a more appropriate type of agnostic attitude is the attitude of the agnostic who is undecided, but ready to stop being undecided upon further considerations. An agnostic with this initial attitude might be swayed toward incompatibilism by manipulation cases because those cases could count as clarifying considerations (as they do count, according to Pereboom, because they draw attention to the fact that one's action is causally determined by factors beyond one's control). Thus, agnosticism about the manipulated agent's moral responsibility does not follow automatically from agnosticism about the compatibility of moral responsibility and ordinary causal determinism.[179]

So, in order to defend the hard-line strategy, a compatibilist has to show that manipulation cases do not count as clarifying considerations. In what follows, I argue that the compatibilist can do that only if he rejects the possibility of moral responsibility for actions that are not performed for the right reasons.

[178] Michael McKenna, "A Hard-line Reply to Pereboom's Four-case Argument," 152-154.
[179] Derk Pereboom, "A Hard-line Reply to the Multiple –Case Manipulation Argument," *Philosophy and Phenomenological Research* 77 (Jul., 2008): 163-164.

5.2.5 The Reason View and the Manipulation Argument(s)

To see why manipulation cases may not be so helpful for understanding the relation between responsibility and determinism, we need to pay attention to an asymmetry in our intuitive reactions to manipulation cases. The asymmetry in question consists in the fact that certain forms of 'ordinary life manipulation', especially those which result in virtuous behavior seem much less troubling than the forms of manipulation cited for the purposes of the manipulation arguments. Consider, for instance, the following case presented by McKenna:

> A young child, let us call her Ann, watches up close the deterioration and death of a parent from a crippling disease, leukemia, medically addressed when treatments like chemotherapy were in their infancy, when they were simply barbaric. Suppose that this child, well before the age of mature reason, and so gripped by such an experience, simply came to see life as limited, precious, but also chocked with the prospects of suffering and tragedy. From this she comes to see her life as one that should not be squandered, that should be lived to its fullest, with no promise of a long future or a lovely afterlife. Whether for good, rational reasons or not, suppose those experiences settled for that child what would become her deepest unsheddable values about how to live. And suppose that as a mature adult she acts upon them. Does she do so unfreely? Is she not responsible for the conduct issuing from those values?[180]

According to McKenna, this case is "very much like a manipulation case, except that the manipulation is not by the design of a team of scientists like Team Plum, but by the vagaries of life." However, as McKenna points out, the agent herself in this case (who is in reality his friend) does not regard her situation "as an impediment of her freedom and her responsibility or... her dignity, but as a condition of it."[181]Nomy Arpaly makes similar observations about cases of people who have undergone radical transformations, "for reasons that were entirely beyond their control" (e.g. transformation from 'party animal' to workaholic, from person

[180] Ibid.156.
[181] Ibid.

who lacks desire for parenting into a loving parent, or various religious conversions).[182]These cases seem to show that manipulation per se is not what drives our intuitions that people in some manipulation cases are not morally responsible. But what is then the explanation of our intuitions in those cases? Why is there such an asymmetry in our intuitions? According to McKenna, the asymmetry is due to the fact that incompatibilists refer to very unusual cases for which our intuitions are not, so to say, well prepared and thus tend to be misleading. McKenna argues for this claim in the following passage:

> Our intuitions have evolved along with our ordinary practices. It is only to be expected that when those intuitions are tested in extremely different contexts, contexts which differ radically from the ones out of which they evolved, they will be indecisive. If we had, as Wittgenstein might have put it, a very different ''form of life'', one where some of us maybe many or even all of us, were presumed to be manipulated by teams like Team Plum, our intuitions might be quite different about these cases.[183]

Pereboom admits that there is an asymmetry in his own intuitions about manipulation cases. For, he admits that his position "has the cost of denying that McKenna's causally determined and perhaps manipulated virtuous agent is morally responsible."[184]In addition, he says that he believes that McKenna's intuition about responsibility of the woman in his example is widespread. But, his explanation of this intuition is that acting virtuously is still something valuable that should be celebrated even if the agent, because of the truth of determinism, "does not deserve, in the basic sense, praise for her efforts."[185] In other words, according to Pereboom, our intuitive reactions to some ordinary manipulation-like cases do not show that our reactions to manipulation cases are misleading because our reactions to those ordinary

[182] Ibid.
[183] Michael McKenna, "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument,"157.
[184] Derk Pereboom, "A Hard-line Reply to the Multiple-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77, No 1 (Jul., 2008), 167.
[185] Ibid.

cases can be explained by pointing to factors which are irrelevant for judgments of moral responsibility.

However, there are also compatibilist explanations of why manipulation seems less troubling when it leads to good actions. The one available to the proponents of the Reason View seems to me the most promising. Thus, according to Dana Nelkin, our intuition that agents who are manipulated to do bad things cannot be morally responsible is due to the fact that manipulation usually deprives agents of the ability to recognize and act for the right reasons. On the other hand, in her view, our mixed feelings about the cases in which manipulated agents do good things for good reasons can be explained by pointing to the fact that in those cases manipulation does not deprive agents of the ability to recognize and act for good reasons (and in fact may provide them with that ability). Therefore, the Reason View enables us to say that manipulation per se does not account for the intuition that agents in some manipulation cases are not responsible, but its association with its likely effect, i.e. the lack of the relevant abilities.

The problem with this explanation, however, is that it does not eliminate *all* worries about manipulation, especially when the result of manipulation is a good action. For, even though one's possession of the relevant abilities could result from manipulation that is not the main effect of manipulation. The main effect of manipulation is determination of what the agent does, i.e. determination of whether he exercises or refrains from exercising his abilities. And since the Reason View on Wolf/Nelkin's interpretation offers no explanation of why the agent finally does what he does which refers to some property of the agent, it is reasonable to conclude that what the agent does is not up to the agent but up to the manipulators. Consequently, the manipulators rather than the agent seem to be responsible for the agents' action.

Now, it is easy to show how rejecting the possibility of the unexercised ability to do the right thing for the right reasons could help with this problem. For, if there is no such ability, it is literally true that manipulators' work in determining agents to perform the right actions for the right reasons consists solely in helping them to acquire the abilities in question. For, once the agent acquires the relevant abilities, he will exercise them and there will be no space for further influence of the manipulators. Furthermore, it makes sense to say then that what the agent finally does is up to him because the explanation of what he does refers to one of his properties, i.e. to his ability to do the right thing for the right reasons.

Therefore, by modifying the Reason View in the way I suggested, we can explain why manipulation or determination by factors over which one has no control does not undermine one's responsibility. So, manipulation arguments give us a reason to accept the view that we can be free and responsible only when we do right things for the right reasons.

5.3 Conclusion

In conclusion, there are (at least) two paths to the view that we can be free and responsible only when we do right things for the right reasons. The first path starts from the acceptance of the Reason View and incompatibilism about ability to do otherwise and determinism. Those who accept these assumptions must take this path if they want to avoid absurdities, that is, they must take it if they want to avoid the conclusion that mere addition of indeterminism could turn someone who lacks free will into someone who has free will. The second path starts from the acceptance of the Reason View and acceptance of compatibilism about the relation between free will and deterministic origins of actions. Accepting these positions leads to the above mentioned conclusion because they cannot otherwise give a plausible answer to the manipulation argument. Thus, it seems that anyone who accepts the

Reason View must also accept the claim that we can be free and responsible only for the right actions performed for the right reasons.

CONCLUSION

My main goal in this dissertation has been to show that free will as a power required for moral responsibility is possible. I have argued that we are in the best position to defend free will if free will is the ability to do the right thing for the right reasons which cannot be unexercised. Put differently, the result of my research is that free will is possible if it is the capacity to understand and act on the basis of the right reasons, and if given the possession of this capacity, it is not possible *not* to act in accordance with the right reasons or to do the right thing. In what follows, I will summarize my reasons for drawing this conclusion.

My main reason for accepting this view is that it follows from the Reason View, which says that free will consists of the ability to do the right thing for the right reason. It follows from this view because without rejecting the possibility of unexercised ability to do the right thing for the right reasons its proponents cannot give an adequate account of self-determination. For, as I argued at the end of chapter 4 and in chapter 5, this is clear when we ask what accounts for the agent's exercise of the ability in question on particular occasions or consider situations in which the actions of the agent who has that ability are results of causal determination or manipulation by another agent. For, unlike the traditional view according to which free will consists in the ability to do otherwise, this view cannot point to a property of the agent which explains why it is up to the agent rather than to pure chance or manipulator how he acts on particular occasions. By rejecting the view that the ability which is central to this view can be unexercised one can explain why the agent acts in a particular way on a particular occasion. One can do that simply by citing the presence or the absence of that ability.

Another reason why I have found this conclusion plausible is that it follows inevitably from the acceptance of the asymmetry of the Reason view and incompatibilism about ability

to do otherwise and determinism - the thesis I accepted in the first chapter on the basis of my consideration of the Consequence Argument. As we have seen, this is so because there is no other way for those who accept these assumptions to avoid the absurd conclusions that the mere presence of chance could make a difference between free and non-free agent or the conclusion that no one ever acted on the basis of the right reasons. However, as I have noticed, this reason is relevant only for those who have not yet recognized that the Reason View must be modified in the way I suggest and who accept incompatibilism and I don't know of anyone who has this combination of views at the moment.

Obviously, the key assumption in my reasoning has been that the Reason View is essentially correct. This view, as we have seen, has clear advantages over other views about free will currently on offer. The main advantages of this view over other 'rationalist' views of free will appear when we consider conditions of responsibility for the wrong actions performed for the wrong reasons. According to this view, in contrast to other such views, to be responsible for actions of this sort, that is, to deserve blame, one needs to be able to do otherwise or to avoid blameworthiness. In this respect, the Reason View is more in line with our common sense understanding of responsibility than those other views. For, in ordinary life we accept the lack of ability to avoid wrongdoing as a valid excuse. It is also in harmony in this respect with the intuitively plausible 'ought implies can' principle which rules that there is no wrongdoing and consequently no blameworthiness without ability to do otherwise because something counts as a wrongdoing only if it is true that one ought to do otherwise or ought to refrain from doing it. The most serious challenge to this aspect of the Reason View is based on the assumption that the so called Frankfurt-style cases in which the agents are blameworthy although they could not do otherwise are possible. However, I believe that in the second chapter I presented good reasons for rejecting the possibility of such cases.

The Reason View also has significant advantages over views which ground free will in some metaphysical condition such as the availability of alternative possibilities or some sort of power to originate one's actions which is incompatible with determinism. Its key advantage over the former view becomes visible when we focus on what I have called the Luther cases in which agents have very good reasons for what they actually do and no good reasons nor a strong inclination to do otherwise. In typical examples of these cases, agents seem to act freely and deserve praise, although it is difficult to see how the ability to do otherwise could ground their control over their actions. For, the exercise of that ability would seem to be rather a failure than exercise of control. The advantage of this view in other cases (in which doing otherwise would not obviously constitute a failure to control) is not so clear, but it is also not clear why we should prefer the traditional view over this view when it comes to those cases. The only reason for that, as my inquiry has shown, could be the necessity of ability to do otherwise for self-determination. But, with the modification of the Reason View that I have suggested, the traditional view loses even this advantage over the Reason View. The same is true of the views according to which free will consists in some sort of power to originate actions which requires indeterminism. For, the only reason one might prefer those views over the Reason View could be that they provide an adequate account of self-determination.

Although the view that I have suggested differs in one significant respect from the Reason View in its standard form defended by Susan Wolf and Dana Nelkin, I believe that it shares with it all the advantages over other views. For, according to the view that I have suggested, sufficient condition for free will required for moral responsibility is the same as the sufficient condition of the Reason View in its standard form: it is the ability to do the right thing for the right reason. It is true, though, that my view accounts for self-determination by postulating the relation of determination of free actions by this ability and

by identifying this ability with the agents true self (which makes it also a sort of Real-Self View). However, this relation is not some additional power but simply a result of the claim that the ability in question cannot be unexercised.

It might seem, though, that this addition is not so harmless when we consider free will and responsibility under assumption of indeterminism. For, it might seem that the Reason View modified in this way requires determinism. And this might seem to some as big disadvantage of the view given the widespread scientific belief that all processes are indeterministic at the level of basic physics. It might seem, in other words, that given our scientific knowledge, according to this view our free will hangs on a dangerously thin thread.

However, I don't think that the view I suggested has this consequence. First, this is so because we can simply stipulate that the ability that is central to the Reason View is necessarily exercised unless indeterminacy at the neural level interferes with it. More plausibly, we can assume that our having of the ability to do the right thing for the right reasons depends on non-actualization of certain possibilities. Luther cases again provide support for this claim. For doing otherwise would be a sign that something went wrong in the agents mind, that is, it would be a sign that the agent either lost his ability to appreciate the right reasons or to act in accordance with them.

The view that I have suggested, of course, seems to have one big disadvantage over the standard version of the Reason View and all other non-skeptical views: it denies free will in cases of doing wrong things for the wrong reasons. But how big a disadvantage this really is? It is certainly a major disadvantage if the primary aim of the theory of free will is to give a theoretical underpinning of commonsense and to justify our current practices. However, it is not clear how reliable commonsense is when it comes to these matters. Moreover, it is not clear what commonsense says when it comes to these issues. Certainly, our commonsense does not says so clearly that we sometimes freely do wrong things as it says that there is

external world. For, when we start thinking about wrong things we have done we can always find some explanation why we did what we did, that is, we always some ordinary factor that may serve as an excuse. It seems that we don't need to invoke evil demons or present dream scenarios to make sense of the claim that we might be regularly wrong in our judgments about freedom and responsibility.

The denial of free action in case of wrongdoings would also be a major disadvantage if the aim of a theory about free will were necessarily a defense of our practices of blaming and punishing people or justification of negative emotional responses such as anger or resentment. But, whether these emotions and practices are something that we must defend is controversial. For, it is not clear that these feelings and practices are really something for which we value free will. After all, wouldn't we be better off without those feelings and practices? Wouldn't the world be a better place if people tried to understand why wrongdoings occur instead of blaming and resenting the wrongdoers? The fact that the positive answers to these questions do not sound implausible explains to some extent the appeal of skepticism about free will and moral responsibility.

So, the view that I have suggested has significant advantages over other non-skeptical views and it is not clear that the fact that it is revisionary with respect to our ordinary judgments about wrongdoings outweighs its advantages. Moreover, we have good reasons to think that its advantages outweigh its disadvantages because we have good reasons to think that accepting this view is the only way to show how it is possible to have free will. Whether this is definitely the case, I must admit, has to wait for the results of further research, especially concerning the nature of Luther cases which play very important role in my argument.

Having said this, I don't have any illusions that this will convince many people accept this view. But, it should be very interesting to skeptics. For, it shows that they are not

entirely wrong about free will. It shows that they are right that all existing views of free will fail. But they are wrong that revisions in our conception of free will have to be so big that it is difficult to see why we discuss free will in the first place. To save free will we might have to change our understanding of free will radically, but changes required may not be so big that they bring into question its significance.

BIBLIOGRAPHY

Arpaly, Nomy. *Unprincipled Virtue*. New York: Oxford, 2003.

Balaguer, Mark - *Free Will as an Open Scientific Problem*, MIT Press, 2010.

Beebee, Helen. „Reply to Huemer on the Consequence Argument." *The Philisophical Review* 111 no.2 (Apr., 2002): 235-241.

Beebee, Helen. Local Miracle Compatibilism. *Noûs* 37 (2003): 258-277.

Berofsky, Bernard. "Review of Susan Wolf's Freedom within Reason", *The Journal of Philosophy*, Vol. 89, No. 4 (Apr., 1992): 202-208.

Berofsky, Bernard. "Global Control and Freedom." *Philosophical Studies* 131 (2006): 419-45.

Carlson, Eric. "Counterexamples to Principle Beta: A Response to Crisp and Warfield," *Philosophy and Phenomenological Research* 66, No. 3 (May, 2003): 734-736.

Chisholm, Roderick "Human Freedom and the Self," in *Free Will*, edited by Robert Kane, Blackwell, 2001.

Clark, Randolph. *Libertarian Accounts of Free Will*. Oxford: Oxford University Press, 2003.

Demetriou, Kristin."The Soft-Line Solution to Pereboom's Four-Case Argument," *Australasian Journal of Philoso*phy 88 (2010): 595-617.

Crisp Thomas M., and Warfield, Ted A. "The Irrelevance of Indeterministic Counterexamples to Principle Beta," *Philosophy and Phenomenological Research* 61, No 1 (Jul., 2000): 173-184.

Dennett C. Daniel. *Elbow Room*: *The Varieties of Free Will Worth Wanting*. Cambridge. MA: MIT Press, 1984.

Double, Richard. "Review of Susan Wolf's Freedom within Reason." *Mind*, New Series 101 No. 401 (Jan., 1992): 198-200.

Double, Richard. *The Non-Reality of Free Will*. New York: Oxford University Press, 1991.

Finch, Alicia. "On behalf of the consequence argument: time, modality, and the nature of free action." *Philosophical Studies* 163 (2013): 151–170.

Fischer, John, Martin and Ravizza, Mark. "Review of Susan Wolf's "Responsibility, Freedom, and Reason Freedom within Reason." *Ethics* 102, No. 2 (Jan., 1992): 368-389.

---------- *The Metaphysics of Free Will*. Oxford: Blackwell, 1994.

---------- and Ravizza, Mark. *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press, 1998.

---------- "Responsibility and Manipulation." *The Journal of Ethics* 8 (2004): 145-77.

---------- "Responsibility and Alternative Possibilities," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006): 27-52.

---------- Kane, Robert, Pereboom, Derk and Vargas, Manuel. *Four Views on Free Will*. Oxford: Blackwell Publishers, 2007.

---------- "'Ought-Implies-Can', Causal Determinism and Moral Responsibility." *Analysis* 63 (Jul., 2003): 244-250.

---------- and Garret Pendergraft "Does the Consequence Argument Beg the Question?" *Philosophical Studies* 166 (2013): 575-595.

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66 (1969): 829-39.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68 (1971): 5-20.

Gallois, Andre. "Van Inwagen on Free Will and Determinism," *Philosophical Studies: An International Journal for Philosophy in Analytic Tradition* 32, No. 1 (Jul., 1977): 107-111.

Ginet, Carl. "In Defense of Incompatibilism." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 44, No. 3 (Nov., 1983): 391-400.

Ginet, Carl. *On Action*. Cambridge: Cambridge University Press, 1990.

Graham, A. Peter. A Defense of Local Miracle Compatibilism.*Philosohical Studies: An International Journal for Philosophy in the Analytic Tradition* 140 (2008): 65-82.

Haji, I. *Moral Appraisability*. New York: Oxford University Press, 1998.

Hunt, David P. "Freedom, Foreknowledge and Frankfurt," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006).

Huoranszki, Ferenc. *Freedom of the Will: A Conditional Analysis*. New York: Routledge, 2011.

Huoranszki, Ferenc. Powers, Dispositions and Counterfactual Conditionals, *Hungarian Philosophical Review*56 (2012): 33-53.

Huemer, Michael. "Van Inwagen's Consequence Argument." *The Philosophical Review* 109, No. 4 (Oct., 2000): 525-544.

Hume, David. *An Enquiry concerning Human Understanding.* Oxford: Oxford University Press 2007.

---------- *A Treatise of Human Nature.* Mineola: Dover 2003.

Kane, Robert. *The Significance of Free Will.* New York: Oxford University Press, 1996.

---------- "Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson." *Philosophy and Phenomenological Research* 60 (2000): 157-167.

--------- ------- *Free Will.* Malden: Blackwell 2002.

---------- "Responsibility, Indeterminism and Frankfurt-style Cases: A Reply to Mele and Robb," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, edited by David Widerker and Michael McKenna. Ashgate, 2006.

Kapitan, Tomis. A Master Argument for Incompatibilism. In The Free Will Handbook, ed. Robert Kane (Oxford: Oxford University Press, 2001): 127-157.

Klein, Martha. *Determinism, Blameworthiness, and Deprivation.* Oxford: Clarendon Press 1990.

Levy, Neil. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press, 2011.

Leibniz, G. W. *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil,* Trans. E. M. Huggard (La Salle, Ill.: Open Court, [1710] 1985), 421.

Lewis, David, "Are We Free to Break the Laws." *Theoria* 47 (1981): 113-21.

Locke, John. *An Essay Concerning Human Understanding,* edited by P.H. Nidditch, Oxford: Clarendon Press, 1689/1975.

Lowe, E. J. *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press, 2008.

McCann, Hugh. *The Works of Agency: On Human Action, Will, and Freedom*. New York: Cornell University Press, 1998.

McKenna, Michael. "A Hard-line Reply to Pereboom's Four-case Argument." *Philosophy and Phenomenological Research* 77 (2008): 142-159.

---------- „Compatibilism, in *Stanford Encyclopedia of Philosophy*, 2009.

----------"Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism." *Journal of Ethics* 16 (2012): 152-153.

Mele, Alfred. ''A Critique of Pereboom's 'four-case argument' for Incompatibilism.'' *Analysis* 65 (2005): 75–81.

---------- *Free Will and Luck.* Oxford: Oxford University Press, 2006.

---------- and David Robb, "Bbs, Magnets and Seesaws: The Metaphysics of Frankfurt-style Cases," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 128.

---------- "Manipulation, Compatibilism, and Moral Responsibility." *Journal of Ethics* 12 (2008): 263-86.

Mickelson, Kristin. “The Zygote Argument is Invalid: Now What?” *Philosophical Studies*172 (2015): 2911-2929.

Mickelson, Kristin. “A Critique of Vihvelin’s Three-Fold Classification,” *Canadian Journal of Philosophy* 45 (2015): 85-99

Narveson, Jan. “Compatibilism Defended.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 32, No 1 (Jul., 1997): 83-87.

Nelkin, Dana, Kay. *Making Sense of Freedom and Responsibility*, Oxford: Oxford University Press, 2011.

Nelkin, Dana Kay. “The Consequence Argument and the Mind Argument.” in *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*, edited by Paul Russel and Oisin Deery, 126-134. Oxford University Press, 2013.

O’Connor, Timothy. *Persons and causes*, Oxford: Oxford University Press, 2000.

O’Connor, Timothy. “Agent-Causal Power,” in *The Philosophy of Free Will*, ed. Paul Russell and Oisin Deery.

Palmer, David. “The Timing Objection to the Frankfurt Cases.” *Erkenntnis*: An International Journal of Scientific Philosophy 78 (2012): 1011-1023.

Pereboom, Derk. "Determinism *Al Dente*," *Noûs* 29, 1995, 21-45.

---------- *Living without Free Will*, Cambridge: Cambridge University Press, 2001.

---------- "Source Incompatibilism and Alternative Possibilities," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006): 185-201.

---------- "A Hard-line reply to the Multiple-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77 (2008): 160-70.

Pink, Thomas. "Power and moral responsibility." *Philosophical Explorations* 12:2, (2009): 127-149.

Rosen, Gideon - "Skepticism about Moral Responsibility", *Philosophical Perspectives*, 18, *Ethics*, 2004.

Russell, Paul. „Hume on Free Will," in *Stanford Encyclopedia of Philosophy* (2007)

Slote, Michael. "Selective Necessity and the Free Will Problem." *Journal of Philosophy* 79 (January, 1982): 5-24.

Smith, Michael and Pettit, Philip. "Freedom in Belief and Desire", in *Mind, Morality and Explanation*, Oxford: Oxford University Press (2004): 375-396.

Strawson, Galen. *Freedom and Belief*. Oxford: Clarendon Press 1986.

---------- "The Bounds of Freedom" in *The Oxford Handbook of Free Will*, ed. Robert Kane (2002).

Strawson, Galen. *Freedom and Belief*, Oxford: Oxford University Press, 1986.

Strawson, Peter F. Freedom and Resentment," Proceedings of the British Academy 48 (1962): 1-25.

Stump, Eleonore. “Moral Responsibility without Alternative Possibilities,” in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006), 140.

Van Inwagen, Peter. „Reply to Narveson.“ *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 32 (1977): 89-98.

---------- *An Essay on Free Will*. Oxford: Clarendon Press, 1983.

---------- “When is the Will Free,” *Philosophical Perspectives* 3 (1989): 409.

----------Free Will Remains a Mystery. Philosophical Perspectives 14 (2000): 1-19

Vihvelin, Kadri. „Arguments for Incompatibilism,“ in *Stanford Encyclopedia of Philosophy,* 2011

----------“The Modal Argument for Incompatibilism," *PhilosophicalStudies* 53 (1988): 227-44.

----------“Free Will Demystified: A Dispositional Account.” *Philosophical Topics* 32 (2004): 427-450.

---------- *Causes, Laws, and Free Will: Why Determinism Doesn’t Matter*. Oxford: Oxford University Press, 2013.

Wallace R. Jay. *Responsibility and the Moral Sentiments*. Harvard University Press, Cambridge, Massachusetts London, 1998.

Watson, Gary. “Free Agency.” *Journal of Philosophy* 72 (Apr., 1975): 205-20.

---------- „Review of Susan Wolf's Freedom within Reason", *The Philosophical Review* 101, No. 4 (Oct., 1992): 890-893.

---------- "Two Faces of Responsibility." *Philosophical Topics*, 24/2 (Fall 1996), 227-48.

---------- "Soft Libertarianism, Hard Compatibilism," in *Agency and Answerability* (Oxford: Clarendon Press, 2004): 197-215.

---------- "Reasons and Responsibility," in *Agency and Answerability: Selected Essays* (Oxford: Clarendon Press 2004), 294-301.

Widerker, David. "Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities," in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Ashgate, 2006)

Wolf, Susan - *Freedom Within Reason*. Oxford: Oxford University Press, 1990.

Zimmerman, Dean. "That Was Then, This Is Now: Personal History vs. Psychological Structure in Compatibilist Theories of Autonomous Agency." *Noûs* 37: 638-71.

----------"Reasons-Responsiveness and Ownership-of-Agency: Fischer and Ravizza's Historicist Theory of Responsibility." *Journal of Ethics* 6 (2003): 199-234.