

PRACTICAL PERSPECTIVE COMPATIBILISM

Sofia Jeppsson



Practical Perspective Compatibilism

Sofia Jeppsson

©Sofia Jeppsson, Stockholm 2012

ISSN 0491-0877

ISBN 978-91-86071-84-4

Printed in Sweden by US-AB, Stockholm 2012

Distributor: Department of Philosophy, Stockholm University

I thank my supervisors Ragnar Ohlsson and Gunnar Björnsson, for their invaluable help. I thank my husband Alexander, who was always there for me during hard times. I thank my loving family, my parents Jan-Erik and Lena Nordin as well as my sisters Anna Nordin and Elin Pellberg, for their support.

Contents

Introduction	9
1. Moral Agency and Moral Responsibility	14
2. Practical Freedom: Historical Background of the Concept	18
2.1. Deliberation and Determinism	19
2.2. Kant on Freedom and Morality	25
2.3. Kierkegaard	29
2.4. Korsgaard	30
2.5. Bok on Moral Responsibility	32
3. The Freedom Sufficient for Morality	38
3.1. The Principle of Deliberative Relevance	38
3.1.1. Morality as Action-guiding	39
3.1.2. Morality from a Second and Third-Person Perspective	45
3.1.3. Moral responsibility and Perspectives	48
3.2. A stipulative Definition of Practical Freedom	51
3.2.1. Minimal and Maximal Practical Freedom	52
3.2.2. Unconsidered Options	58
3.2.3. More or Less Practical Freedom	67
3.2.3. Diminished Practical Freedom Compared to Weakness of Will	69
3.2.4. Free Choice of Values	75
3.2.5. The Things We Just Do	77
3.2.6. The Importance of Being Able to Act on One's Own Values	79
3.3. Practical Perspective Compatibilism Compared to Causal Compatibilism	83
4. Different Perspectives	88
4.1. Different Perspectives – Contradictory Beliefs?	88
4.2. A contextualist Interpretation	95
4.2.1. Lewis-style Contextualism	95
4.2.2. Williams-style Contextualism	102
4.2.3. Contextualism about the Moral Words	109
5. Moral Responsibility	112
5.1. Desert-entailing Moral Responsibility	112
5.2. Formal Arguments for Incompatibilism	118
5.2.1. A General Problem with these Arguments	118

5.2.2. Galen Strawson's Argument for the Impossibility of Moral Responsibility	119
5.2.3. The Consequence Argument and Frankfurt Responses	127
5.3. Thought experiments for incompatibilism	140
5.3.1. Hardline and Softline Replies	141
5.3.3. A Softline Reply for Vehicle Scenarios	142
5.3.4. A Hardline Reply for Walden Two	144
5.3.5. A Hardline Reply for Professor Plum	148
5.3.6. A Hardline Reply for Alice	152
5.4. Arguments for Causal Compatibilism	160
5.4.1. A Hardline Reply for Mele's Manipulation Cases	161
5.4.2. Kaye's Mitigating Factors	167
5.5. Summary of Chapter 5	173
6. Moral Agency, Responsibility and Metaethics	175
6.1. What Determinism Can and Cannot Imply	175
6.2. Korsgaard and Constructivism	176
6.3. Contractualism	180
6.4. Peter Strawson's Naturalism	183
6.5. Realism	186
7. Conclusion	193
Bibliography	201

Introduction

It can only be true that agents acted morally right or wrong if they acted freely. This statement is rather uncontroversial. However, any particular interpretation of “freely” will be controversial. Some philosophers believe that we cannot be free if the universe is deterministic, others that we cannot be free if it is indeterministic, and some think the morally relevant kind of freedom is impossible regardless of how the universe works. Moral philosophy thus has a problem; we discuss the right- and wrong-making features of actions and talk about people as being responsible for what they do, but it is not clear if there can *be* such things as rightness, wrongness and moral responsibility in the world. The purpose of this dissertation is to solve this problem.

The free will problem is an old one; it has now haunted philosophy for a couple of thousand years. The problem has changed shape over the centuries, but the enduring core of the problem is this: How can I act freely, if something outside me (be it the laws of logic, the laws of nature, or the almighty will of God) determines everything that will happen, including my actions? *Determinism*, as the term is used in the contemporary debate, means that the past and the laws of nature together determine everything that will happen in the future. Nevertheless, if that is so, how can I be free? Or to put it a little more stringently; suppose that P is a proposition describing the state of the entire world at some arbitrary point in the distant past, and suppose that L is a proposition describing all the laws of nature. Proposition F describes a future action of mine. Now if P and L imply F, is it not the case that F *had* to

happen? Was I not, in that case, powerless to falsify F, powerless to do anything but the action that F describes?

Other philosophers have contested this and argued that determinism is no threat to freedom; some have even argued that determinism is necessary for freedom. They argue that I am free when I do what I want, when my actions are caused by my desires. If my desires did not determine what I do, these doings would seem like random events rather than free actions. The philosophers who hold this view are called (classic) compatibilists, while their opponents are called incompatibilists. When it comes to free will and determinism, incompatibilists may believe either that we can be free if the world is *indeterministic* in the right way or that the only alternative to determinism is a regrettable sort of randomness. Incompatibilists who believe in free will are called libertarians (not to be confused with the political group of the same name). There are different libertarian theories of what makes us free. Some think it is a little bit of quantum indeterminacy at the right time and place in our brains, while others think it is agent causation, a special kind of causation whereby agents can initiate completely new causal chains.

Almost all philosophers who have written on the subject fall rather neatly into one of the two camps; compatibilist or incompatibilist. Either they believe that the right kind of deterministic causation is no threat to freedom, and may even be a necessary prerequisite for freedom, or they believe that determinism and freedom are mutually exclusive. There is an interesting exception to this rule: Immanuel Kant. Most philosophers place him in the incompatibilist camp, some in the compatibilist, but I think Ted Honderich is right when he argues that Kant does not fit nicely into either one. He calls Kant “an incompatibilist of a unique kind” in *How Free Are You*¹, and a “metaphysical compatibilist” in *On Determinism and Freedom*.² I will argue

¹ Honderich, 2002, p 5

² Honderich, 2005, p 150

in this dissertation for a Kant-inspired theory on freedom and morality which is compatibilist, but different from most theories in that camp.

I will argue that *practical freedom*, the sense in which one is free when one has to decide what to do (even if that decision in turn were determined by the past and the laws of nature), is all the freedom one needs for moral agency and responsibility. Questions of what kind of freedom one needs are roughly speaking questions of in what sense, if any, agents need to have alternative possibilities open to them and in what sense, if any, they need to be the origin of their actions. I will argue that moral agency requires the ability to choose between actions that are open to agents in the sense that, as far as they know, they can perform them if they try to. Moral agency also requires that agents are the origin of their actions in the sense that *they* must choose their actions, they cannot wait for the past and the laws of nature to choose *for* them. Various capacities that have nothing to do with freedom might also be required for moral agency. Perhaps, for instance, only agents who are capable of forming relationships with other creatures can be moral agents. However, in this essay I will merely argue that practical freedom is sufficient freedom-wise for moral agency; what other capacities an agent might need to be a moral agent will not be investigated.

I will argue that practical freedom is sufficient freedom-wise because morality is action-guiding. What makes an action morally right or wrong must be factors that either make action-guiding possible in the first place by making choice and deliberation possible, or directly guide action by figuring as reasons for choosing one option rather than another. Having the power of agent-causation, for instance, is not necessary for deliberation and choice, nor can it figure as a reason for choosing one option rather than another. Therefore, the power of agent causation cannot be a condition for one's actions being right or wrong. I will also argue that moral responsibility is plausibly tied to rightness and wrongness, so that the same freedom conditions hold for the former as for the latter. I call the thesis that practical freedom

suffices, freedom-wise, for moral agency and responsibility “practical perspective compatibilism”, or PPC for short.

PRACTICAL PERSPECTIVE COMPATIBILISM: The thesis that practical freedom suffices, freedom-wise, for moral agency and moral responsibility.

In order to fully appreciate my thesis, it is of course necessary to have a firm grasp of the concepts moral agency, moral responsibility and practical freedom. The first chapters of this dissertation are intended to provide that. In chapter 1 I explain in some detail what I mean by “moral agent”, and how that concept relates to that of moral responsibility. Chapter 2 provides the reader with some historical background to the concept practical freedom, and the thesis that it suffices for moral agency and moral responsibility. Chapter 3 begins with an argument for the sufficiency of practical freedom: Morality is action-guiding, meaning it is supposed to be used in deliberation and advice. Therefore, factors that are irrelevant for deliberation and advice are also irrelevant when making moral judgements. This conclusion I call “the principle of deliberative relevance”, or PDR for short.

THE PRINCIPLE OF DELIBERATIVE RELEVANCE: Only factors that are relevant for a deliberator or adviser are relevant when judging whether somebody did right or wrong, or was morally responsible for his actions.

I use PDR as an argument for PPC. Having some kind of freedom other than practical freedom is irrelevant for deliberation and advice, and therefore irrelevant when making moral judgements. If we can judge whether somebody’s action was right or wrong, or whether they were morally responsible, without taking into account whether they were free in more than the practical sense, it follows that practical freedom suffices freedom-wise for moral agency and responsibility. Thus PDR provides an argument for PPC. Then I

provide the reader with definitions of maximal and minimal practical freedom, the factors that enhance and diminish this kind of freedom, and how moral agency and moral responsibility co-vary with practical freedom. Both PDR and the practical freedom definitions build on the theories discussed in chapter 2. They differ by being neutral between different ethical theories (while many of the philosophers from Chapter 2 tie their theory of freedom to Kantian ethics) and by allowing for degrees of practical freedom, since the philosophers from Chapter 2 usually discuss freedom as something one either has or has not.

Chapter 4 deals with some concerns regarding the Kantian idea that we can view the world from different perspectives. Practical freedom is the freedom we have from a practical perspective, the perspective of deciding what to do. From a theoretical perspective, the perspective of explanation and prediction, we may seem completely determined or just random. One concern is that the theoretical perspective shows us the truth, and the “freedom” we think we have when deciding what to do is really an illusion. I explain how one can give a contextualist interpretation of this talk of different perspectives. Which factors are relevant change from context to context. We are free in the context of deciding what to do, because determining factors and quantum randomness cease to be relevant in this context.

I give an argument *for* PPC in chapter 3. Chapter 5 deals with famous arguments *against* PPC, or against groups of theories where PPC is included, and show why they all fail. Chapter 6, finally, shows that metaethical considerations further support compatibilism, since incompatibilism implies the falsity of a number of respected metaethical theories.

1. Moral Agency and Moral Responsibility

The free will debate traditionally focuses on moral responsibility, not moral agency. It has been argued that one cannot be responsible for what one did if one could not have done otherwise, and determinism precludes the ability to do otherwise. It has also been argued that one cannot be responsible for an action unless one was responsible for what caused this action, and for what caused the cause of the action etc. Whether the universe is deterministic or not this responsibility condition creates an infinite regress impossible to satisfy. This traditional focus on moral responsibility may lead people to believe that only those who subscribe to a normative-ethical theory that employs this concept need concern themselves with questions regarding free will. This is not so.

Suppose first that determinism threatens moral responsibility because it makes all actions inevitable, so that no one can ever do anything but what he or she actually does. Suppose also that Kant was right when he claimed that ought implies can. These two premises taken together imply that it is never the case that agents ought to do something other than what they actually do. If we assume that agents do wrong only when it is the case that they ought to do something else, it also follows that there can be no wrongdoing in a deterministic world. If rightdoing requires the possibility of wrongdoing, it follows that there is no rightdoing neither. These problems must be addressed by moral philosophers even if they do not employ the concept of moral responsibility. As Ishtiyaque Haji put it; determinism is a threat to

“the moral anchors”, the concepts of right, wrong and ought.³ However, what about indeterminism? If indeterminism is a threat towards moral responsibility, because one cannot be responsible for what happens by chance, it seems to be an equally serious threat towards right- and wrong-doing, since we do not call that which happens by chance either right or wrong, but fortunate or unfortunate. Since I do not want to give the false impression that the free will problem becomes irrelevant if the true normative-ethical theory is one that does not employ the concept of moral responsibility, I will formulate my primary thesis in terms of moral agency, the capacity to do right or wrong. However, in Section 3.1. and Chapter 5 I will discuss the notion of moral responsibility, and explain why I do not think it requires anything more freedom-wise than moral agency.

The following seems to me to be the most plausible account of the relationship between moral responsibility and moral agency: Moral agency is necessary but not sufficient for moral responsibility, while moral responsibility is sufficient but not necessary for moral agency. Moral responsibility requires first that there are moral agents, and second that the true or the best moral theory is one that employs the concept of moral responsibility. Suppose, for instance, that utilitarianism is the true normative ethical theory. This theory does not employ the concept of (desert-entailing) moral responsibility. So if utilitarianism is true, nobody is morally responsible in this sense, for reasons that might have nothing to do with free will or determinism. Since it could still be true that agents perform actions that are either morally right or morally wrong, there could still be moral agents. Derk Pereboom, who is not a utilitarian, also has a moral theory according to which moral agency is possible even if moral responsibility is not.⁴ On the other hand, Haji argues that it is possible to be morally responsible in a blamewor-

³ Haji, 1999, p 175

⁴ Kane, ed, 2002, p 479

thy sense, even if no actions are right or wrong.⁵ However, I find this very counter-intuitive.

I will argue that two conditions, the existence of moral agency and that the true or correct moral theory employs the concept of moral responsibility, are both necessary and jointly sufficient for the existence of moral responsibility in the world. I will argue that determinism and indeterminism are both compatible with moral agency as well as moral responsibility. However, I never argue that there *is* moral responsibility in the world – I leave it open whether this concept belongs in the correct or true ethical theory or not.

I wrote that a moral agent is an agent who can do right or wrong. However, why choose a disjunctive rather than conjunctive formulation? Suppose there is an individual who is extremely motivated to always do what is morally right, never makes any mistakes figuring out what the right thing to do is, and always successfully acts on his motivation; let us call him an angel. In a sense, he *cannot* do wrong. Given his psychology the probability that he will perform a morally wrong act is zero, or so close to zero that it makes no difference. Now “can” is an ambiguous word, and there might be some sense of the word according to which it is possible for this individual to do wrong, as long as he is in control of himself and as long as his devotion to rightness cannot be described as a neurosis that compels him to do right. However, since he at least *arguably* cannot do wrong, and I still would like to label him a moral agent, I choose a disjunctive rather than conjunctive formulation. The same goes for a devil, who is completely devoted to doing what is morally wrong all the time. We may imagine that he is non-neurotic, in control of himself, just completely devoted to evil. Arguably he cannot do right

⁵ Haji, 1999, p 190. He thinks that Frankfurt has shown that one can be morally responsible for inevitable actions. However, he also thinks that since ought implies can and wrongdoing implies that one ought to have done something else instead, an inevitable action cannot be *wrong*.

given his particular psychology, but I still do not think one should define a moral agent so that it becomes true by definition that the devil is not one.

It is important, however, that moral agents inhabit a world where both right- and wrongdoing is possible. If the world is such that no agents, regardless of their psychology, could do wrong, morality loses its action-guiding function.⁶ The same problem arises if the world is such that rightdoing is impossible. Therefore, the world must be such that both right- and wrongdoing are possible in it if there is to be any moral agency.

⁶ In Section 3.1.1. I will write more about what it means that morality is action-guiding.

2. Practical Freedom: Historical Background of the Concept

Throughout the history of the free will debate there have been incompatibilists who argue that there could not be morally responsible agents, moral agents or even rationally deliberating agents⁷ if we lived in a deterministic universe. Some philosophers, beginning with Kant, have chosen a particular approach when responding to these claims. They have argued that the perspective we view things from when making choices and deciding what to do is the one relevant for questions of whether it is rational to deliberate or whether actions can be right and wrong. From this perspective we often do have several options open to us however deterministic the universe may be. In this chapter I will describe their arguments as well as how they view freedom and its connection to rational deliberation and morality. In the next chapter I will lay down a definition and an argument that attempts to capture the core of what these philosophers have to say on the subject.

⁷ For instance van Inwagen, 1983, p 160, Ginet, 1990, p 90, Haji, 2009 p 169. Aristotle, 1998, pp 55-58 [1112a11-1113a6] can also be read this way and is sometimes regarded as the first philosopher to advocate agent causation. On page 56 he writes: "Deliberation is concerned with things that happen in a certain way for the most part, but in which the event is obscure, and with things in which it is indeterminate". For my own part, I do not think it obvious that Aristotle's thoughts on deliberation make it irrational to deliberate about my own actions under known determinism as long as I do not know what will follow in this particular instance.

2.1. Deliberation and Determinism

Tomis Kapitan has argued in “Deliberation and the Presumption of Open Alternatives” that known determinism poses no threat to rational deliberation. The philosophers that Kapitan addresses in his article, like Hector-Henri Castaneda and Richard Taylor, argue roughly as follows: I cannot rationally deliberate about whether to bring about some event that I think is determined to happen.⁸ Suppose, for example, that I think it is determined that the sun will rise tomorrow – it would then be irrational for me to deliberate about whether to prevent tomorrow’s sunrise. Since it is irrational to deliberate about what one believes to be determined, all deliberation would be irrational for somebody who believed that *everything* is determined.⁹ If we were to find out that determinism is true, we must either give up deliberation, or, since this is probably impossible, be doomed to irrationality – so the argument goes.¹⁰

Kapitan, on the other hand, argues that rational deliberation that is about whether to A only requires that my A-ing as well as my not A-ing is contingent relative to the beliefs I hold at the time, and also that I believe I will A or not-A as I decide to. The first condition he calls *the presumption of contingency*,¹¹ the latter *the efficacy condition*.¹² Tomorrow’s sunrise is safely outside the scope of deliberation since it fails to meet both these conditions. Firstly, my beliefs about astronomy imply that the sun will rise tomorrow, so it is not contingent relative to my beliefs. Secondly, I do not think there is any connection between my decisions and sunrises. The sun will rise regardless of what I decide. Since Kapitan has a compatibilist explanation of why

⁸ Kapitan, 1986, p 231

⁹ *ibid* pp 230-231

¹⁰ van Inwagen writes (1983, p 160) that “to reject free will is to condemn oneself to a life of perpetual logical inconsistency”. Free will, as van Inwagen here uses the word, means that the past and the laws of nature do not necessitate a particular action on my part. van Inwagen thus wants to argue that somebody who believes his actions to be determined by the past and the laws of nature, is condemned to a life of perpetual logical inconsistency.

¹¹ Kapitan, 1986, pp 233-237

¹² *ibid*, pp 232-233

we cannot rationally deliberate about such things as preventing tomorrow's sunrise, incompatibilists about determinism and rational deliberation cannot use these kind of examples to support their thesis. Events that *can* be rationally deliberated about according to Kapitan's theory are easy to find. Here is a simple everyday example: Suppose I ask myself if I should have my coffee break now, or in half an hour. As I ask myself this I have many beliefs, but none of them implies the exact time of my coffee break. I may have the *general* belief that everything is determined, but since I lack detailed beliefs about the past and the laws of nature, my general belief in determinism cannot entail anything in particular about my coffee break. Both having my coffee break now and having it in half an hour is therefore contingent relative to the beliefs I hold at this time, including my belief in determinism. I also believe that if I decide to have my coffee break now I will, and if I decide to have it in half an hour I will. This is also consistent with my belief in determinism, since it is a conditional statement.¹³ Thus, I satisfy all Kapitan's conditions for being able to rationally deliberate about my coffee break.

To show that an efficacious will and contingency relative to my other beliefs, rather than some fancier kind of freedom to have my coffee now or abstain from it, is sufficient for rational deliberation, we may toy with the idea that it is only rational to deliberate about actions that we could agent-cause. Immediately we run into difficulties, the first one being that I have no idea whether I have this power or not. I would thus have no idea whether I could rationally deliberate at all. Suppose now that God reveals that I do not, in fact, have the power to agent-cause things. What should I then do about

¹³ In a sense, the statement "If I decide to prevent tomorrow's sunrise, I will prevent tomorrow's sunrise" is also true under determinism, as long as I do not decide to do this. A material implication is true if the antecedent is false. To make Kapitan's thesis plausible one should rather interpret the fact that I will have my coffee break in half an hour if I decide to have it in half an hour and will have it now if I decide to have it now, as a statement about nearby possible worlds, or something along those lines. In this world I took a coffee break right now, but

my coffee break, since I cannot rationally deliberate about it? Should I just sit here and wait to see what will happen, when my coffee break will appear? If I do that, I will certainly go without coffee – at least unless a colleague takes pity on me in my philosophical paralysis and brings me some. But why would that be more rational than simply deciding to have coffee and then having some? As soon as I have my epistemic openness and efficacious will it is simply necessary for me to deliberate in order to get things done, and it is hard to see how it could then be irrational to do so.

The efficacy condition may remind readers of Moore’s classic “conditional analysis” of the word “can” – that somebody can do something if she would do it if she chose to – and the objections launched against it. Chisholm has argued, for instance, that I do not really have the ability to do something just because I could do it if I chose to, since I may be unable to make the choice in the first place.¹⁴ Neurotic people, for example, may seem unfree even though they satisfy the conditional, because they cannot choose what to do.¹⁵ Now I do not think this is true about all neurotics. Some neurotic persons may be able to choose to go against their neurosis, but fail to satisfy the efficacy condition. Suppose that I have an extreme fear of the dark. I hate to have this phobia, so I want to go out into the dark and face my fear in order to overcome it. I could choose to do so, but then I would start to tremble in the doorway and fail to take another step. My fear is simply so extreme that I cannot control myself in the face of it. I can thus make the choice, but will fail to act on it. However, in Section 3.2.2, I deal with cases where people are unable to make a certain choice because they cannot seriously deliberate about that option. In my terminology, agents may satisfy Kapitan’s efficacy condition and still fall short of *maximal* practical freedom, because there are

there’s a nearby possible world where I waited half an hour. However, any world where I prevent tomorrow’s sunrise must be extremely different from the actual one.

¹⁴ Chisholm, 1976, p 57

¹⁵ See, for instance, O’Connor, 2010, subsection 1.2.

some choices they cannot really deliberate about. I will explore this subject in depth later in the dissertation.

Susan Wolf, on the other hand, has argued that the idea that one will do A if one decides to A, and refrain from A if one decides to refrain from it, is *too strong* as an analysis of general ability.¹⁶ I have the ability to understand English, even if I cannot decide not to understand it. However, as a condition of what it requires to deliberate about an option, rather than general ability, it works. I cannot deliberate about whether to understand an English-speaking person, precisely because I will understand what he says regardless of my decisions. I *can* deliberate about whether to engage in conversation with him, because I will either do or not do that as I decide to.

Kapitan concludes that the feeling of freedom we can experience in choice is the feeling of epistemic openness and our wills being efficacious.¹⁷ An agent with epistemic openness and an efficacious will has practical freedom.¹⁸

Kapitan thinks that having practical freedom is necessary and sufficient for rational deliberation, and one has practical freedom when the efficacy condition and the presumption of contingency are fulfilled. However, some philosophers contradict Kapitan when it comes to the presumption of contingency, claiming that this is not necessary for practical freedom and rational deliberation at all. Christine Korsgaard suggests as much in her *Sources of Normativity*.¹⁹ Likewise Dana Nelkin explicitly argues that the efficacious will is all we need.²⁰ Hilary Bok on the other hand, has given an argument to the effect that we never *could* have certain knowledge of our future actions before we decide upon them anyway. Suppose, Bok writes, that determinism

¹⁶Wolf, 1990, p 99

¹⁷ Kapitan, 1986, p 232

¹⁸ *ibid* p 241

¹⁹ Korsgaard, 1996a, pp 94-96. She confirmed in personal correspondence, June 20 2008, that this is indeed what she means.

²⁰ Campbell, ed, 2004, p 108

is true, and there was a device called the Pocket Oracle (PO) that could calculate my every move before I made it from facts about the past and the laws of nature. It might be perfect as long as I did not look at the predictions, but once I did, PO would run into an infinite regress. Since my looking at the predictions would affect my behaviour, PO could not predict my behaviour unless it knew how its prediction would affect me. To know that, it would have to know what its prediction was going to be before it made the prediction – and that is impossible.²¹ I do not believe this argument is sound. Although my looking at PO's prediction would affect my behaviour, determinism means that PO could calculate everything that will happen based on, for example, the state of the universe in 1982 and the laws of nature. The state of the universe in 1982 and the laws of nature would imply *both* PO's prediction *and* my reaction to that prediction, so PO would in fact not have to know its own prediction before it could know anything about my reaction. However, I suspect there may be counter-arguments to my argument here. I will therefore not attempt to settle the matter once and for all. Let us just, for the sake of argument, assume that PO was invented, that it *was* impossible to disobey, and I had one. Or let us suppose that I became perfectly clairvoyant. Would this mean that I could not or needed not deliberate and choose any longer? To return to the coffee break example, suppose that I can predict with absolute certainty that I will have my coffee break in half an hour from now on. Can there still be some kind of *choice* involved? Do I still *decide* to have my coffee break in half an hour rather than right away? I have no idea whether the answer to that question is yes or no. Perhaps prediction of my own actions could somehow replace deliberation and choice if prediction were perfect. Perhaps I would still have to deliberate and choose in order to act, despite knowing in advance how my reasoning and subsequent choice would play out. If I still needed to deliberate in order to act, and deliberated about actions where my will was efficacious (rather than, say, tomorrow's

²¹ Bok, 1998, pp 81-87

sunrise), it seems that it would still be rational to do so. However, I do not think this question is particularly important. PO does not seem to be in any danger of being invented, and there is no sign of the human species evolving perfect clairvoyance. Whether it is a contingent or a necessary truth that all actual agents who deliberate and choose their actions also have an epistemically open future is not relevant for my project. If real people in the real world can be rational deliberators, this suffices for the purpose of this dissertation.

What about indeterminism then? It could perhaps be argued that indeterminism means that things happen by chance, and it is irrational to deliberate about whether to bring something about, if that something will or will not happen depends on brute chance. I can deliberate whether to take the lottery ticket on the right, but I cannot deliberate whether to take the winning ticket, since whether I win depends on brute chance. But even if the world contains some random elements, there are still lots of fairly reliable connections between events. Known indeterminism could only destroy Kapitan's efficacy condition if it meant that there was really no connection between decisions and actions, but that seems implausible and is contradicted by countless everyday experiences. For all we know it might be the case that indeterministic quantum events have some large scale effects that produces truly random events. However, even if physics would one day prove this, it could still be the case that under almost all circumstances I do what I have decided to do.

From now on, I will regard it as established that deliberation about what to do can be rational regardless of (known) determinism or indeterminism. We often find ourselves in situations of choice. Should I have my coffee break now or in half an hour? Whether the world is deterministic or indeterministic, it can still be the case that I will have it now if I decide to have it now or in half an hour if I decide to wait that long. It might be the case that I will not-A if I decide to not-A and A if I decide to A. This is enough for

deliberation to be rational. I need not presuppose that the future is open in some absolute sense in order to rationally deliberate about what to do, as long as I am confident that I will act on my decisions. Next I will take a look at philosophers who argue that some kind of practical freedom, the ability to consciously choose one's action and having an efficacious will, also suffices for moral agency and moral responsibility.

2.2. Kant on Freedom and Morality

Immanuel Kant famously distinguished between theoretical and practical reason. Reason is, indeed, only one faculty, but it can be employed in two different ways. In its theoretical use, it is something we use in trying to figure out how things *are*. In its practical use, it is something we use in trying to figure out what to *do*. We grasp how things are when we learn more and more about the laws of nature that determine the world around us. We decide what to do when we regard all the options before us and then settle on one of them. This distinction between practical and theoretical reason is the acknowledged inspiration for Kapitan's argument for the sufficiency of practical freedom for rational deliberation. According to some interpretations, Kant claimed that practical freedom was also sufficient for moral agency, since it is sufficient for deliberation, and deliberation requires morality.²²

Kant started to develop this idea in *the Foundation of Metaphysics of Morals*. His aim was to show that agents with free will must be moral, by showing that the law of a free will is also the moral law.²³ Since he wanted this argument to apply to us humans, not merely some hypothetical free-willed beings, he had to argue first that we have free will.

Kant writes that since we have reason, we do not just *follow* laws, but reflect over and consider laws. We do not just find ourselves doing things, but

²² Korsgaard and Bok, whose views I will describe in more detail later, interpret Kant in this way.

²³ A rational "must", I think.

consider what to do. This is what it means to say that we have practical reason.²⁴ Practical reason means deciding what to do. When I decide what to do, I cannot think of my will as being decided by something else – so Kant claims.²⁵ One may object to this that many philosophers believe in determinism, and think their wills are “decided by something else”, the past and the laws of nature, while still being able to decide what to do. Kant was aware of this, so it is plausible to assume that he did *not* mean that it is impossible to make up one’s mind if one is a determinist. He might have meant, since people often have inconsistent ideas, that all determinists who still manage to decide what to do also believe that their decisions are undetermined whenever they deliberate – as van Inwagen, Castaneda, Taylor and others have claimed. However, I think the most plausible interpretation is that focusing on *causes* that determine my will one way or the other would just amount to introspection. No decision will appear if one introspects instead of deliberating. To decide something one must focus on the alternatives one chooses between rather than the causes of one’s will – one must think about possible options rather than determining causes. Kant also writes that I *necessarily* conceive of myself as free when I consider what to do. Once again, I think it is implausible to suppose that this means that I, if I normally believe in determinism, necessarily cease to do this whenever I deliberate. A better interpretation is that I necessarily focus on alternatives rather than causes when choosing what to do, since otherwise it would not *be* choosing but rather an attempt to predict my own behaviour (which would always end up in me doing nothing, if I just sat there and introspected trying to figure out what I will be caused to do next). I might find out, through introspection, that I feel most inclined at the moment to watch TV, but this will not lead to action unless I decide to do what I feel most inclined to do (after all, it happens that I feel most inclined to watch TV yet sit down to work). Thus, I conceive of

²⁴ Kant, 1997 pp 37-38, and 2004 pp 695-696

²⁵ *ibid* pp 76-77

myself as free when I think of various actions as alternatives that are open to me, because I can perform them if I decide to. Kant has thus argued that I must think of myself as free when deciding what to do; given this interpretation I think it is a valid claim.

When I conceive of myself as free, I must also act like a free person.²⁶ Kant claims that freedom can be neither mere randomness, nor being decided over by something or somebody else, so a free will must make its own law to follow.²⁷ It has a law, so it is not random, but it made the law itself, so it is not decided over by something or somebody else – it is truly free. If I must conceive of myself as free, and if “free” means making my own law, I must also make my own law to act on – otherwise I would be inconsistent. However, if I both think *as if* I were free, and act *as if* I were free, then there is *no* discernible difference between me and an imagined free person – which means that for all practical purposes, I *am* free.²⁸

It may seem as if a step in the argument were skipped here, but I disagree. Picture two agents, one with practical freedom in a deterministic universe, and another with some kind of libertarian freedom that means there is truly a garden of forking paths in front of him. They *both* must make up their minds, they *both* must decide what to do, and they *both* need to base their choices on some reason or principle since choices cannot be completely random. So where is the difference between them? Perhaps, the difference is that the second agent’s options have some kind of existence that the first agent’s options lack. From some metaphysical or perhaps merely physical perspective one can point out differences between these agents, but if *being* an agent with practical freedom is exactly like *being* a libertarian agent, then from a practical perspective there is no difference to be found; both are equally free.

²⁶ I think this “must” is best understood as rational necessity, the necessity of not having contradictory thoughts.

²⁷ *ibid* 76-77

²⁸ *ibid* 76-77

Kant thought that when taking up a theoretical perspective (for example a physical one) in a scientific context we would always find determining causes (perhaps not immediately, but as science progressed we would find determining causes for more and more phenomena). That would be the case regardless of what one chooses to study, be it human beings or something else.²⁹ Still, the perspective we view things from in action, when deciding what to do, and the perspective we view things from in science (and the proto-science we constantly do in everyday life), when we try to find out how things are, are on equal footing in Kant – or if anything has primacy it is practical perspective.³⁰ Both perspectives are absolutely necessary to us; neither is more true than the other one. They simply play different roles.

Kant then thought that a person with practical freedom is necessarily a moral agent. This is where the claim that a free will must make its own law comes in. A person who must choose his actions must have some kind of principle or maxim from which he can conclude that one action is better than the other. In today's philosophy it is more common to talk about reasons, but saying that a person must have reasons to be able to choose his actions (rather than just pick them at random) amounts to more or less the same thing. If I take a pill for the reason that it will cure my illness, another way to express the same thing is to say that I act on the maxim "when ill, if there is a medicine that can cure you, take said medicine" (or something similar). Kant then thought that his own system of morality was the only one that could be rationally justified and explained all the way down, and that it was therefore not only necessary for people with practical freedom to have *some* principle, but also rationally necessary for them to adhere to Kantian morality. Kantian morality, summed up in the categorical imperative, would thus simultaneously be "the law of a free will" and the "moral law". The categorical imper-

²⁹ Kant, 2001 pp 124-129, and 2004 p 693

ative imposes no restrictions on the will except that it should follow principles – which the will must do if it is to be a *will* at all rather than a disjointed bundle of desires pulling this way and that – but still a number of substantial duties can be inferred from it (at least according to Kant).

Even if one disagrees with Kant about the possibility of inferring substantial moral principles from the mere necessity of conforming to a law, one can agree that he has shown that there is a connection between practical freedom and morality. Having practical freedom *makes* us think in normative and moral terms. Creatures lacking practical freedom can simply go where their urges drive them, but in order to *choose* what to do, we must rank options as good or bad, better or worse, to consider some facts reason-giving and others not. Simply liking some things and disliking others is not enough, since we may like something and yet question whether we should pursue it.

In the practical perspective the past and the laws of nature become irrelevant, I am the relevant cause of my action, the one responsible for bringing it about. These factors are not contingently true about agents with practical freedom, in the way that it is contingently true about most Swedes that they speak Swedish (one can easily imagine a situation where Swedish vanished and was replaced by some other language). Instead, these factors follow from the very fact of having practical freedom – *choosing* one's actions rather than grabbing one option at random *means* having some kind of norms, principles, reasons, as well as seeing oneself as responsible for bringing the action about.

2.3. Kierkegaard

An existentialist like Kierkegaard argues for a similar conception of freedom. His famous statement that life must be lived forwards but understood

³⁰ See for example Kant 2004 p 698 on the three basic questions of reason, and 2001 p 88 about the laws of the will compared to the laws of nature.

backwards is reminiscent of the Kantian distinction between theoretical and practical reasoning, and freedom as something residing in the practical perspective.³¹ When studying former actions one may understand their causes, but actions that still lie before us must be *chosen*. We can study history and see how the actions of large groups of individuals happen by necessity, by law. To think that free will makes a science of history impossible is to misunderstand the subject-matter.³² Still, the individual has his freedom to choose between various alternatives open to him. Kierkegaard's aesthete writes that he is like a spider, hanging from a thread. Behind him are causes driving him forward, but in front of him there is just a vast openness.³³ Kierkegaard's ethical man also stresses the necessity, the inevitability, of choice. The ethical man describes the aesthete as somebody who does not truly choose, but allows himself to be swayed this way and that by urges and impulses. Still, even the aesthete makes a choice of sorts. He chooses to be the *kind of person* who is swayed this way and that by urges and impulses. In this weak sense of choice, everybody must choose.³⁴ The connection between practical freedom and ethics is that true choices, where one actively decides what to do, are the subject-matter of ethics. The ethical man has a higher kind of freedom in that he continuously chooses what to do, while the aesthete just choose the general principle "I will let myself be swayed by my impulses and urges".

2.4. Korsgaard

Christine Korsgaard is a contemporary Kantian philosopher who has further developed Kant's theory of practical freedom. For Korsgaard, practical free-

³¹ Kierkegaard, 1968, fourth volume, note 164

³² Kierkegaard, 2002b, p 168

³³ Kierkegaard, 2002a, pp 27-28

³⁴ Kierkegaard, 2002b, pp 161-164

dom arises from self-awareness.³⁵ A less intelligent animal is aware of its surroundings and has a limited self-awareness in that it knows where it is positioned relative to those same surroundings. Yet, desires are simply something that drives the animal to act. Hunger drives it to search out food, fear of something drives it to escape, and sometimes contradictory desires give rise to inner conflict. Although it experiences desires, it is not aware of them *as* desires. It cannot consider the fact that experiences of hunger, thirst and a need for companionship are all various desires, and call into question whether to pursue them or not on a given occasion.³⁶ Practical freedom comes from having this awareness. Instead of simply having a desire rising up in me which then pulls me towards its object, I become aware of having a desire, and must then decide whether to act on it or not. Here, we once again arrive at the view that morality naturally belongs to the practical perspective. When deciding whether to pursue a certain desire or not, I must judge it to be either a good or a bad one, that it either promotes some purpose I consider valuable or on the contrary is detrimental to some important value.³⁷ Without normative judgement I would not have anything to base my decision on. This is how practical freedom brings normativity and morality with it.

The description of Korsgaard's theory might remind some readers of Harry Frankfurt's. Frankfurt claims that just as a free action is one where the agent does what he wants to do, a will is free when the agent wants what he wants to want. When his first-order desires (desires relating to things that are not desires) are in line with his second-order desires (desires relating to first-order desires), he is free.³⁸ This is sometimes referred to as a hierarchical perspective on freedom. One might think that Korsgaard's view is hierarchical as well, since there is an agent over-and-above the various desires that must decide whether he likes them or not. However, on Korsgaard's view

³⁵ Korsgaard 1996a pp 92-93

³⁶ Korsgaard 2008, p 4

³⁷ *ibid* p 94

³⁸ Frankfurt, 1988, pp 168-176

there is a difference between desires on the one hand and the agent on the other hand. Desires are pulls, urges and mere feelings, and these can be thought about and reflected upon. The one doing the thinking and reflecting is not himself another desire, but an agent.³⁹ This picture is compatible with the claim that this same agent, from some kind of psychological and scientific standpoint, can be described solely in terms of beliefs and desires of various orders.⁴⁰ From the practical perspective he must still decide whether to endorse and act or resist various desires that he finds in himself, so from the practical perspective he still cannot be reduced to a bundle of desires. Just as an agent cannot sit back and wait for the past and the laws of nature to cause his actions, he cannot sit back and wait for the desires to act *for* him. He can of course decide to sit back and wait for something to happen, but in that case he has already made a decision, and if he wants an end to his wait he better make another one. In the sense relevant to the practical perspective, it is the agent that is responsible for bringing the actions about, not some cause further down a causal chain stretching backwards in time. Factors further down the causal chain may be interesting and explanatorily important from a theoretical standpoint, but when deciding what to do one cannot delegate the choices to them.

2.5. Bok on Moral Responsibility

Bok is another contemporary Kantian who makes an explicitly compatibilist interpretation of the theory of practical freedom. Unlike Korsgaard – who focus primarily on duties, that is, the forward-looking consideration of what one ought to do – Bok aims to explain why practical freedom is sufficient for backward-looking moral responsibility.

³⁹ See for example Korsgaard, 1996a pp 230-231

⁴⁰ *ibid* p 101

Bok argues that many compatibilist accounts simply fail to address the real concerns that are voiced by incompatibilists. She coins the term *mechanism* for the view that all events in the world are metaphysically similar in important respects. All events, including human actions, may either be determined links in long causal chains or be subject to the same kind of random quantum factors. However, whether determinism is true or not, there is no deep metaphysical difference between human actions and everything else.⁴¹ When we consider why something happened, human choices may figure in our explanation of the event, just like many factors that are not human choices, but there is nothing metaphysically special about choices that validates talk about “moral responsibility”.⁴² If I am morally responsible for killing a man, but an avalanche cannot be morally responsible for killing another man, we want to know what the relevant difference is. The murder I committed may have been caused by my feelings and thoughts, while the avalanche may have been caused by some meteorological and geological factors. That certainly is *a* difference, and no incompatibilist would deny that. However, it is not clear why that difference is *morally relevant*, if feelings and thoughts as well as meteorological and geological factors are part of the same causal net of events, or subject to the same kind of random quantum events. Saying that I could have done otherwise if I had wanted to, if my thoughts and feelings had been a little bit different, does not address the issue either, since the avalanche could have slid otherwise if the geological and meteorological conditions had been a little different.⁴³ Neither does Bok think that the fact that it is possible to influence me by blame is obviously morally relevant. There may be various ways to influence my future behaviour, just like there may be various ways to prevent future avalanches. Perhaps one way to influence my future behaviour is to act towards me *as if* I were morally responsible, but that does not prove that I *am*. One can imagine

⁴¹ Bok, 1998, p 3

⁴² *ibid* pp 34-35

some non-person that could be influenced by the words “you’re bad!” but which is certainly not morally responsible – Bok’s example is a computer program that were designed to change its behaviour whenever one typed the words “bad program”.⁴⁴ This fact about the program obviously does not make it morally responsible. The fact that people’s future behaviour can be affected by praise and blame does nothing to prove that people can be morally responsible for what they do.

The morally relevant difference between the causes of my own behaviour and the causes of things that happen is, according to Bok, the following: When engaged in deliberation, I must disregard the causes of my own behaviour and instead focus on my reasons for action.⁴⁵ If I turn my attention to a phenomenon like an avalanche, however, there is no alternative to focusing on the causes. Human actions and avalanches may be metaphysically similar when we are engaged in theoretical reasoning, trying to give causal or scientific explanations of an event, but only actions can be regarded from a practical perspective. When I am engaged in deliberation and deciding what to do, it is necessary for me to view myself and my actions from the practical perspective.

When it comes to people other than me, I have a choice. When I deliberate, I could either regard them as *things* or as other agents. However, Bok thinks the former is ethically wrong; to respect people is to think of them as agents like me, to view them from a practical perspective⁴⁶. To think of them as agents is to focus on their reasons for action rather than on causes of their behaviour.

To be engaged in practical deliberation is to ask myself what I should or ought to do in this situation. The use of some kind of forward-looking oughts

⁴³ *ibid* p 102

⁴⁴ *ibid* pp 34-35

⁴⁵ *ibid* pp 104-109

⁴⁶ *ibid* pp 189-194

and should⁴⁷ is therefore justified by our need for deliberation and choosing our actions. The justification of backward-looking ascriptions of responsibility is a little bit trickier to explain. Bok's idea is that I am blameworthy for a bad act if I did it because my will failed, and praiseworthy if my will prevailed and I did something good. If I did something yesterday that I *now* judge to be wrong, but I sincerely believed yesterday that it was the right thing to do (perhaps I now have more information on the matter than I had yesterday), I will not blame myself, although I could regret what I did. However, if I let myself give in to laziness instead of doing what I actually knew was best, if I let myself be deluded by a self-serving bias instead of thinking things through properly, if my will failed to transfer my best judgement on what I should do into action, I should blame myself. Blame directed towards myself is simply the judgement that I have a faulty will that showed itself in that bad action of yesterday. Thinking that there is a fault in my will is, from a practical perspective, different from thinking about any other fault. Firstly, a faulty will does not just affect what alternatives I have to choose between or which alternative is the best one, but my ability to choose among the alternatives I do have. If I become severely ill or handicapped (Bok's example is if my lungs failed and I had to stay in an oxygen tent at all times) my options would be drastically limited, but with a well-functioning will I could still perform the best one of the few options that remain.⁴⁸ If my will falters, it might not matter if I have lots of good alternatives to choose between since I might pick a bad alternative anyway. A faulty body may mean that I cannot achieve my original goals and have to lower my aims (perhaps drastically), but a faulty will means that I violate my own principles and fail to achieve some goal that is perfectly available to me. Secondly, my will is what all other practical interests I have depend on. All other practical interests lie in the pursuit of ends that I have adopted by will. If my will is flawed, I will not

⁴⁷ For a discussion of different kinds of ought, see Section 3.1.1.

⁴⁸ *ibid* p 148

adopt the right ends in the first place, but may aim for goals that I would, if I had reflected properly on the matter, have rejected. More particular practical interests could be given up or changed if something about my body or my circumstances make them difficult or impossible to achieve, but since the will is what all other practical interests depend on I cannot give up the interest I have in a well-functioning will. If my will is at fault, I must blame myself and try to change. This is why it is justified to hold ourselves morally responsible – from the practical perspective my will is crucially different from anything else in nature. Holding others responsible is justified because I should view them like agents, just like me, rather than viewing them like things whose behaviour I try to affect.

I do not think Bok's account is unassailable. She does a nice job of explaining a crucial difference between on the one hand having a faulty will, on the other hand having some other kind of fault in one's body or soul that affects one's ability to do the right thing. But does this necessarily mean that I must *blame* myself when my will falters? Perhaps I could regret that my will faltered and try to improve it without exactly directing blame towards myself. There are, after all, a number of arguments for the incompatibility of moral responsibility and determinism. Perhaps, if I were convinced by these arguments, I could get rid of the impulse to blame myself, while still recognising that my will is crucial for action in a way that other capacities are not and needs to change if it is faulty. There is also something odd about stressing a faulty will as the only justification for blame; surely one can merit blame for an egoistic act, even if one is a convinced egoist and did what one judged was best to do?

When it comes to Kant, Kierkegaard and Korsgaard, they may have successfully shown that there is a connection between practical freedom and morality or at least normativity, in that an agent with practical freedom must think of actions as better or worse in order to choose between them (or at

least, if Kierkegaard was right, if he is to make active choices over and over again rather than just make a one-time choice to follow his spontaneous impulses from now on). Then Kant and Korsgaard actually try to *infer* an entire moral system from the mere fact that we are agents who must choose what to do. If successful, they would have shown not just that practical freedom suffices for moral agency, but also that the true moral theory is the Kantian one. I have not discussed the purported inference of morality from practical freedom in any detail, but suffice to say that most philosophers are sceptical to the idea. I think there is some truth here, but no unassailable argument has yet been given.

In the next chapter I will provide an argument for the sufficiency of practical freedom for moral agency and responsibility. The argument is inspired by the Kantian idea that it is from the practical perspective we view actions as right and wrong, good and bad, but it is neutral between different ethical theories, and does not attempt to actually *infer* morality from freedom. I do not claim that it is a knock-down argument, but together with the fact that all arguments for incompatibilism or causal compatibilism about moral responsibility and determinism fail (something I will show in Chapter 5), I do think it suffices to establish a practical perspective compatibilist conclusion.

3. The Freedom Sufficient for Morality

In this chapter, I will firstly give an argument for the sufficiency of practical freedom for moral agency, secondly define stipulatively minimum and maximal practical freedom, and thirdly explain when agents have their practical freedom compromised and how that affects their moral agency. The previously discussed philosophers tend to view practical freedom as an all-or-nothing affair, either one has practical freedom or not. There are, however, many situations where it seems implausible that people are either perfectly free or not free at all, as in the classic Milgram experiments, in Frankfurt scenarios and when people are in the grip of a neurosis or addiction. I will show that one can allow for degrees of freedom and corresponding degrees of moral agency and responsibility, without abandoning the practical perspective.

3.1. The Principle of Deliberative Relevance

Kant and Korsgaard begin with a theory of freedom, from which they attempt to infer a particular moral theory that has both ethical and metaethical components. They do not discuss whether other moral theories might require some different kind of free will. My aim in this dissertation is different. I want to argue that practical freedom is all the free will one needs for moral agency, regardless of which ethical or metaethical theory is the true one. I argue for this by invoking the principle of deliberative relevance, or PDR for short.

THE PRINCIPLE OF DELIBERATIVE RELEVANCE: Only factors that are relevant for a deliberator or adviser are relevant when judging whether someone did right or wrong, or was morally responsible for his actions.

I argue in turn for this principle by appealing to the fact that morality is action-guiding. It is morality's job to be used by deliberators and advisers. I will first discuss what is and is not relevant for a deliberating agent, and then move on to an adviser, and finally a judge viewing the situation from a third-person-perspective. Although some factors may be either relevant or irrelevant depending on one's theory of reasons and oughts and on whether we take the first-, second- or third-person-perspective, I will show that certain facts (like the universe being or not being deterministic) that are often considered crucial in the debate on free will and morality remain irrelevant throughout.

3.1.1. Morality as Action-guiding

It is often claimed that morality is action-guiding. As Erik Carlson has pointed out, this statement can be interpreted in different ways. It might mean that moral considerations are able to motivate people to act in certain ways, or merely that they tell an agent how to act in a situation.⁴⁹ I will follow Carlson and use the latter definition of action-guiding. A strong argument for the claim that morality is action-guiding is that the concepts of right and wrong seems superfluous if it is not. If morality were merely about evaluation, goodness and badness would suffice. Anybody who agrees that right and wrong are important moral concepts should thus agree that morality is action-guiding.⁵⁰

Carlson has a list of conditions that need to be fulfilled in order for moral theory M to be action-guiding for person P in situation S:

⁴⁹ Carlson, 2002, p 72

1. M must prescribe an action available to P.
2. M must not yield incompatible prescriptions.
3. It is possible for P to deliberate about what to do in S.
4. P is able to understand M.
5. P possesses the necessary empirical information to do what M describes.
6. P is able to make the necessary inferences in a sufficiently short time.⁵¹

I think several of these conditions should go if one tries to make a theoretically neutral definition of what it means for a moral theory to be action-guiding. For instance, can a theory not be action-guiding if it allows for genuine moral dilemmas? It is obvious that a moral theory can guide action in most situations even if it allows for the occasional dilemma, and it can perhaps even be said to be action-guiding in *some* sense in the dilemma situation. If a theory tells me that whatever I do in a certain situation will be wrong, this might mean that I should first pick an action at random and then feel an appropriate amount of guilt, and perhaps perform further actions where I somehow try to atone for what I have done. Something similar could be said for the condition that morality must not prescribe incompatible actions. When it comes to the conditions that P must possess all relevant empirical information and be able to make the necessary inferences, these contradict certain theories on moral reasons and oughts. Some philosophers discuss different kinds of “ought”, where an agent subjectively ought to do what is best given the information he actually has, while he objectively ought to do what is best given everything there is to know about the situation he finds himself in and the consequences of his action.⁵² Other philosophers think all moral oughts are of the objective kind. These philosophers can still agree that morality is action-guiding for agents, even if it is the case that real people in the real world never have enough empirical information or are able

⁵⁰ *ibid* p 72

⁵¹ *ibid* p 73

to make correct inferences regarding what they ought to do. Even if P lacks some necessary information for inferring what M prescribes in S, it could be the case that a hypothetical adviser with more information could use M to give P advice about what he ought to do in S. In that case, I would say that M still is action-guiding for P in S, despite P himself being unable to use it. Morality is not just about deciding what to do, but is also about directing advice and demands towards others, or judging what one would have done in somebody else's shoes. However, all of these aspects of morality are plausibly counted as "action-guiding" rather than merely evaluative. I think that conditions 3 and 4 are sufficient for making M action-guiding for P. As long as P is capable of understanding M and can deliberate about what to do, M could at least provide a hypothetical adviser with something to say to P.

Depending on M, various factors could decide whether an action is right or wrong. For instance, if utilitarianism is part of M, the consequences of actions will figure as right- and wrong-making features. If incompatibilism about right and wrong is part of M, then indeterminism of some kind is both a right-making feature of all right actions and a wrong-making feature of all wrong ones. But what function could an indeterministic principle perform in a theory that is supposed to guide action? If we take the action-guiding nature of morality seriously, all right- and wrong-making features of actions should be such that they are *relevant* for a deliberating or advising agent. This relevance condition means that there can only be right- and wrong-making features of two kinds. Either they make deliberation and conscious choice possible in the first place, or they can figure as reasons *in* deliberation for choosing one action over the other. The first kind of feature makes it possible for the second kind to guide action directly.

Practical freedom in general is a condition of the first kind. Without practical freedom, there can be no deliberation and choice. I have hinted in various places throughout this paper that there are probably conditions for moral

⁵² See for instance Hanson, 2010.

agency besides practical freedom, like the ability to relate to others and see them as reason-giving. If an action can only be right or wrong given that the agent has this ability, it would also be a condition of the first kind. It might be the case that one cannot deliberate about *moral* matters without this ability, and it might therefore be argued that this ability is a right- and wrong-making feature of the first kind. However, neither determinism nor indeterminism is necessary for deliberation and choice, as I have already argued in Section 2.1. Whether determinism or indeterminism is true, it is often both rational and necessary for me to deliberate about various actions that I can perform if I decide to.

The second kind of right- and wrong-making features are such as the consequences of an action, which principles the agent adhered to when acting, which virtues or vices the action manifested etc. Which of these really do make an action right or wrong depend on which normative-ethical theory is the true one. I will not take sides on that matter. However, consequences, moral principles, virtues and vices have the following in common; they can all function as reasons for choosing one action over another. This function makes them crucially unlike determinism and indeterminism. If I have two actions that I must choose between, the fact that whichever action I choose will be determined or undetermined cannot be a reason to pick one of them rather than another. That determinism or indeterminism is true *in general* may figure as a reason for choosing a certain action in special cases; if I know that the universe is indeterministic and somebody asks me whether it is, the fact that it is indeterministic gives me reason to say “yes”. However, even in this situation it is true that the fact my answer being “yes” would be undetermined did not give me reason to do anything in particular.

The lack of practical freedom to perform a certain action is also a reason not to try and perform it. Practical freedom regarding particular actions can thus figure as a reason in deliberation, while practical freedom in general makes deliberation possible. Practical freedom is thus doubly relevant for a

deliberating as well as an advising agent. Yet, determinism and indeterminism, however one looks at it, are not.

I thus think that PDR, the principle according to which only what is relevant to a deliberators and advisers is morally relevant, is supported by the fact that morality is action-guiding. PDR implies that all incompatibilist and causal compatibilist theories are false. If they are false, it follows that practical freedom is all the freedom one needs for moral agency – it follows that PPC is true.

To sum up the argument:

1. (Premise) Morality is action-guiding. We have the concepts of right and wrong because we use morality in deliberation and advice.
2. (Premise, made plausible by 1) The principle of deliberative relevance, which tells us that only factors that are relevant for a deliberator or adviser are relevant when judging whether somebody did right or wrong. Factors relevant for a deliberator or adviser are those that either make deliberation and advice possible or impossible in the first place, or count as reasons for picking one option over another.
3. (Premise) If incompatibilism about right and wrong is true, indeterminism is a right-making feature of all right actions and a wrong-making feature of all wrong actions, and is thus relevant when judging whether an action was right or wrong.
4. (Premise) Indeterminism cannot figure as a reason for or against picking a certain option, and deliberation and choice are possible both with and without it.
5. Therefore (from 2 and 4), indeterminism cannot be one of the features that make right actions right and wrong actions wrong.
6. Therefore (from 3 and 5), incompatibilism about right and wrong is false.

Premise 2 does not follow logically from premise 1, but I think 1 makes 2 highly plausible. Somebody who believes 1 but denies 2 would have to be-

lieve the following: Morality is action-guiding. Therefore, any moral theory M is supposed to be used in deliberation and advice. However, if we are going to use M in deliberation or advice, some factors that have got nothing to do with either deliberation or advice must be in place first. This may not be illogical, but it certainly seems odd.

An argument similar to the one summed up above can be made against the thesis that *determinism* would be necessary for rightness and wrongness, or against a causal compatibilist thesis claiming that an agent must have the right kind of personal history in order to be a moral agent. The right kind of personal history does not make action-guiding possible in the first place, and it does not directly guide action by figuring as a reason for or against various actions. It has nothing to do with action-guidance, and should therefore be considered irrelevant.

Just as I will not take sides in the normative-ethical debate, so too I will not take sides in the debate on reasons and oughts. Some philosophers argue that reasons must be capable of motivating an agent, either as he is or as an idealised version of himself. Others think that reasons for action need no such connection to the agent's motivational state.⁵³ Some think oughts and reasons are relative to the agent's state of information, while others deny this. Even philosophers who agree that, say, the consequences of an action are what determines its moral status, may thus disagree about which consequences perform this function. Is the action made right or wrong by the consequences the agent can be made to care about, the consequences he would care about if he were perfectly rational, the consequences he could expect given the information he had, the consequences he would expect if he had had more information, or all consequences that would in fact follow if he did this or that? The same kind of disagreement can arise between philosophers who are all deontologists or all virtue-ethicists but who have different theories on

reasons and/or oughts. Fortunately, I need not take sides here. The important thing for my purposes is that some considerations will *not* figure as reasons for action regardless of what kind of reasons or oughts we are discussing. If I had the power of agent causation, had been manipulated to acquire the values I now hold, lived in a deterministic universe or had some quantum indeterminacy in my brain, matters that may be interesting from a theoretical perspective, this would not normally give me reason to choose A over B or the other way around.⁵⁴ Thus, if we accept that morality is essentially action-guiding, questions of determinism, indeterminism, agent-causation, manipulation and so on will be irrelevant when judging whether an action was right or wrong, regardless of our theory of reasons.

3.1.2. Morality from a Second and Third-Person Perspective

Morality is not only about deciding for oneself what one ought to do, but also about directing advice and demands towards others, and judging what others have already done. Now if one has a theory about moral reasons and oughts according to which the right thing to do depends either on how much information one has or one's motivational state, the right thing to do from the second-person perspective of a moral adviser might be different from the right thing to do from the first-person perspective of a deliberator. Suppose, for instance, that Dina sits next to a ticking bomb in a building, in a locked room where no one else can get in. The bomb has a red and a green wire. Dina knows that cutting the right wire will disarm the bomb, but cutting

⁵³ See for instance Parfit, 2011b, pp 270-275

⁵⁴ Ragnar Ohlsson suggested that if I discover that I was manipulated to have the values I now have, that could give me reason to seek therapy. I would say that having psychological problems gives one reason to seek therapy, not manipulation in itself, but this is not important. I could grant that people who have been exposed to manipulation, thereby, had reason to seek therapy even if they do not experience any psychological problems. It would not affect my thesis in any important way. Even if it is the case that previous manipulation gives me reason to seek therapy, the fact that the therapy-seeking decision itself will be made by a manipulated agent does not give me reason either for or against it. In that way, manipulation is still irrelevant to the choice at hand.

either the wrong one or none at all will explode the bomb and kill everybody in the building. However, she does not know which wire to cut though. Celine knows about Dina's situation, and she also knows that it is cutting the green wire that will disarm the bomb. She cannot enter the locked room, but she can call Dina on her mobile phone. According to some theories about ought, Dina subjectively ought to cut one of the wires at random, while objectively she ought to cut the green wire. From Celine's subjective point of view it is also the case that Dina ought to cut the green wire, and this is what Celine ought to advise her by phone. If Celine does call her and tell her that she ought to cut the green wire, this will now be what Dina subjectively ought to do as well. The fact that it is the green wire that will disarm the bomb is irrelevant when judging what is subjectively right to do for Dina *as long as Dina does not know this*, but relevant when judging objective rightness or rightness from Celine's point of view, and this fact becomes relevant for Dina's subjective rightness as soon as Dina is given this information. We can add to the situation that Dina is severely depressed to the point where she considers both her own life and the lives of others completely meaningless. According to some theories that tie ought and reason to the agent's motivational state, it is not the case that Dina ought to cut any wire at all, she simply has no reason to do this. It could still be the case that Celine ought to call Dina and do everything she can to talk her into cutting the green wire. The fact that Dina has no wish to live may be relevant when judging what Dina ought to do from her own first-person-perspective, but from the perspective of Celine the adviser it is irrelevant. What is relevant or irrelevant for the question of what one ought to do according to this kind of theory thus depends on whether we view things from the perspective of Dina the deliberator or Celine the adviser. The important thing for my argument is this: It remains irrelevant throughout whether the universe is deterministic or indeterministic, whether Dina has the power to agent-cause things, whether she has been manipulated into valuing or not valuing life, etc. Therefore, these

factors are also irrelevant when judging whether somebody did something right or wrong.

Somebody might object that it would be highly relevant to Celine if she knew that Dina were *determined* to cut a specific wire. This is only true given certain unlikely assumptions. If Celine knew that Dina would cut a certain wire regardless of what anybody told her, giving advice to her becomes pointless. However, determinism does not mean that people in general cannot be affected by what other people tell them. If Celine merely knew that the world was deterministic, and whatever Dina chooses to do will be determined by the past and the laws of nature, she has the same reasons as before to tell Dina to cut the green one. We may discuss in what sense Celine would have a reason to tell Dina to cut the green one if Celine was omniscient, and knew that both her telling Dina to cut the green wire and Dina's subsequent disarming of the bomb were determined by the past and the laws of nature. Yet, as mentioned earlier, I have no idea whether omniscient agents would still in some sense need to deliberate and choose, so I leave that question outside the scope of this essay. Setting omniscience aside, as long as Dina can be affected by advice, determinism will be irrelevant to Celine. Regardless of how the universe works, Celine has reason to call Dina and advise her to cut the green wire.

We may further imagine Lisa, who was not involved in the incident, but read about it in the newspaper the next day. Unfortunately, Celine's mobile phone malfunctioned, Dina cut the red wire, and the building exploded. Lisa may think to herself that Dina did the right thing in the situation, by just picking a wire at random – that way, there was at least a fifty percent chance that she would save the day. Alternatively, she may think that Dina did something wrong unknowingly, or performed an act of blameless wrongdoing. It all depends on Lisa's exact views about wrongdoing. If one has a subjective view on right and wrong, it makes sense to place oneself in the agent's shoes and judge what one would have done (or perhaps would have

done with more moral motivation, if one were more rational, more virtuous etc) given the agent's limited information. If one has an objective view on right and wrong one does not need to place oneself under such constraints. Still, I want to argue that a third-person judgement must be such that it could, hypothetically, have been a piece of advice. This is not to say that Lisa must actively imagine herself advising Dina, only that her judgement should be such that it *could* have been a piece of advice. If Lisa makes the third person judgement that Dina did wrong, this suggests that she would have told Dina not to cut the red wire had she been there (even if Lisa does not actually fantasise about being there). We might think of special cases where one would in fact not have advised what one thinks is right – for instance, we might imagine a psycho killer who considers murdering one person, but would become outraged and kill ten if I dared tell him that murder is wrong. Here we have a situation where there are special reasons to avoid giving *honest* moral advice. It is still the case that hypothetically, if I had been in a position to give some *honest* advice with no disastrous consequences, I would say that murder is wrong. It is simply not the case that the property of wrongness varies depending on whether one is advising or merely judging. Therefore, what is relevant for a third person judgement are the same kind of considerations that are relevant for an adviser – meaning that, once again, determinism and details of Dina's personal history become irrelevant.

3.1.3. Moral responsibility and Perspectives

However, what about moral responsibility? (Let us here, for the sake of argument, suppose that the best normative-ethical theory is one that employs this concept and that, for instance, utilitarianism is false.) Hypothetically, one could judge somebody to be morally responsible for an action from a theoretical perspective rather than a practical one. We could decide to view right and wrong as action-guiding concepts, while seeing moral responsibil-

ity as something that depends on an action's causal history. The question of moral responsibility would then be completely disconnected from the question of right- and wrongness, while the concepts of praise and blameworthiness would consist of a combination of responsibility and right- or wrongness. We would first decide whether agents did right by considering factors that are relevant for choice and advice, as described above. Suppose we conclude that they *did* right. Then we change perspective and consider the causal chain that produced their actions instead. If and only if the causal chain is of the right kind we conclude that they deserve to be praised. But why take this extra step? Why not just praise them because they did what was right? This is, after all, what we normally do in real life. If Sam saves a child from drowning in a lake we will praise him because he did something great, without bothering with some investigation of his genetics and upbringing or ask a physicist to investigate whether there is any quantum indeterminacy in his synapses. The same goes for wrong actions and blame. If Sam had instead pushed the child into the lake we think he deserves blame because he did something wrong. Furthermore, praise and blame are often construed as hypothetical advice. We blame people by telling them that they should have done some other action instead ("Why did you not just leave her alone?"), or pointing out reasons for not doing the action they in fact did ("You could have killed her!"). We praise people by citing factors that counted as reasons for doing the act ("She would have been dead if you had not pulled her up!"). We do, in fact, regard people from the same practical perspective when we praise and blame them as when we advise them or merely judge that they did right or wrong.

Now it is often pointed out that the judgement we pass on Sam for pushing a child into a lake may be mitigated by considering, say, that he was beaten by his parents every day as a kid. However, not everyone will consider this a mitigating factor. Some people will point out that there are plenty of people in the world who have been beaten every day and yet do not push

children into lakes (or something equally bad), so it is not the case that being beaten *cause* one to do so. There is no excuse for Sam not behaving as well as some other people who were beaten by their parents do. Yet, suppose for the sake of argument that we do find Sam's tough childhood a mitigating circumstance that makes him less blameworthy. This can easily be explained by taking up Sam's own perspective and imagining ourselves in Sam's situation. He might be full of rage, which makes behaving decently towards others very difficult for him. He may be impulsive, and not deliberate properly before taking action. That Sam had a tough childhood cannot figure as a reason pro or con pushing children into lakes, and is therefore not relevant when deciding what to do. However, the very process of deciding what to do may be difficult for Sam because of his rage and impulsiveness. Behaving decently may come easy for us, but we may be uncertain about whether we would manage to make the right choices if we were in Sam's situation. I will discuss mitigating factors at length later in this dissertation, but at this point it suffices to say that the mere existence of mitigating factors provides no reason to think that moral responsibility should be assessed by considering the causal history of an action, rather than by considering hypothetical choices made in the agent's situation or hypothetical advice to the agent. We do say that one should not judge anyone until one has walked a mile in his shoes (we shouldn't judge Sam unless we can appreciate what it would be like to choose and act in Sam's situation and with Sam's personality); we do not say that one should not judge anyone until one has investigated the causal chains leading up to his action.

R. Jay Wallace has also argued that when we completely excuse agents who performed seemingly wrong actions, it is because they actually did nothing wrong, or did not do it intentionally.⁵⁵ Blame is justified only when somebody intentionally did something that violates a moral principle. If Anna steps on Billy's foot because Anna was pushed and lost her balance, Anna

did not hurt Billy intentionally. If Anna was threatened by a severe punishment unless she stepped on Billy's foot, Anna (at least arguably) did nothing wrong in saving herself. There are mitigating circumstances where we think people are less blameworthy because their circumstances made it *difficult* to do the right thing, and people are likewise more praiseworthy for doing the right thing when that was difficult. Yet, when we consider people not blameworthy *at all*, it is because they did not intentionally do anything wrong. We place ourselves in the agent's position, and if given her information it would have been right to A, if as far as she knew she had no other option than A, if A was something that merely happened to her rather than an action she chose to perform (like being pushed into stepping on somebody else's foot), or if A was genuinely irresistible and she therefore had no choice, we will not blame her for A.

In the light of all this, it is *prima facie* most plausible to judge moral responsibility from a practical perspective, as tied to right and wrong, and consider the same kind of factors relevant as we do when we judge some action to be right or wrong. The person who claims otherwise, and argues that moral responsibility should be assessed from a theoretical perspective and requires the right kind of causal history behind the action, should be able to back up his claim with a good argument. However, later in this essay I will discuss different kinds of incompatibilist arguments and show why they all fail when directed towards practical perspective compatibilism. I therefore conclude that PDR provides a strong argument for the claim that practical freedom is all the freedom we need both for moral agency and responsibility.

3.2. A stipulative Definition of Practical Freedom

This section will provide definitions of the practical freedom that is necessary for moral agency and also sufficient freedom-wise. When these defini-

⁵⁵ Wallace, 1994, p 135

tions are in place, I will consistently use the term “practical perspective compatibilism” or PPC as a name for the thesis that practical freedom *so defined* is sufficient, freedom-wise, for moral agency and responsibility. Some circumstances and some normative-ethical theories may be more demanding than others, and require close to *maximal* practical freedom in order for agents to count as a full-blown moral agents, or to have full moral responsibility for what they do. According to PPC, however, it is never the case that a different *kind* of freedom is required. Now there are probably other capacities agents must have to be moral agents besides being free – presumably they need the ability to relate to other individuals, and to regard others as reason-giving for them. Exactly what these other capacities are, I will not go into in this essay. It suffices to say that I argue that practical freedom is all the *freedom* they need.

3.2.1. Minimal and Maximal Practical Freedom

Here is the first part of my definition of practical freedom:

MINIMAL PRACTICAL FREEDOM: Agents have minimal practical freedom if and only if they must choose what to do, and they believe their will to be efficacious in realising the considered options they choose between.

Agents with sufficient self-awareness and intelligence to recognise their desires as desires can also ask themselves if it would be good or bad to act on a given desire. If they do not suffer from some strange and irresistible neurosis that compels them to always act on their strongest desire (where “strongest” is defined in a way that does not make it trivially true that everyone always acts out his or her strongest desire), if they are not rendered completely apathetic by severe depression or catatonia, and if they believe

their wills are efficacious, they must choose their actions.⁵⁶ This is both sufficient and necessary for minimal practical freedom.

Their considered options are those alternative actions they actually think about. If some action they could have done had they chosen to never enters their thoughts, it is not true in any sense that they *must* choose whether to do this or not. As I use the word choose, there is really nothing agents can do to avoid choosing in a situation where they consider several options. If they throw a die to decide, it was firstly a choice to throw the die, and secondly another choice to do what the die tells them to. If they hire somebody else to decide for them, the same line of reasoning applies. Firstly they chose to resolve the situation by hiring the decider, secondly they chose to abide by the decider's decision. If they decide to do nothing, this just goes to show that "do nothing" was in fact one of the options they chose between, and by doing nothing they have thus chosen one of the options facing them and performed it.

Practical freedom according to the above definition could be understood both as a general characteristic of an individual, and as something agents can have or lack in particular situations. If agents are, to borrow an example from Aristotle, simply carried along by a strong wind, they are not there and then practically free to choose where to go.⁵⁷ It can still be true that they are the kind of persons who have practical freedom in general. They could still have a level of self-awareness and intelligence that makes choosing what to do rather than just following their instincts necessary in many situations where they are not literally carried away by something or somebody. If so, they are, according to my theory, moral agents in general, although they cannot be said to do either right or wrong in *this* situation.

⁵⁶ I leave "epistemic openness" out of the definition. Arguably, it is implied by "must choose", but as I have mentioned before, some philosophers think I would still have to make choices if I were suddenly granted omniscience, and I will not take sides in that discussion.

⁵⁷ Aristotle, 1998, p 48

The reason for putting a “believe” in the efficacy condition is that if *unknownst* to me I cannot succeed in my attempt unless I try to perform action A, while the seemingly possible actions B and C are really completely impossible, I will still be in a situation where I must choose between A, B and C. The “believe” implies that when two people find themselves in a situation where A is something they can do if they try, while they would fail if they tried B or C, one may have practical freedom to choose between three options and the other one not simply because the first one lacks crucial information regarding the impossibility of B and C. In this particular situation then, practical freedom does not seem worth wanting. It seems better to *know* when one’s will is efficacious than to be mistaken about this matter; better to just choose between actions one could actually perform if one tried to than to spend mental effort on choosing between things one cannot actually do. Even minimal practical freedom based on a false belief about the efficacy of one’s will is morally relevant, however, which the following example will illustrate: There is an election to be held between the Good Party and the Evil Party. The evil neurosurgeon Black wants both Jones and Smith (until now undecided) to vote for Evil. To ensure this, he installs a small device in each person’s brain. The devices will remain mere monitors of the men’s thoughts unless one of them is on the verge of deciding to vote for Good (or stay home at election day) – if that happens, the device will become active and compel the victim to vote Evil. Now suppose Smith finds out that he has this kind of device in his head, and that he is helpless to do anything about it, while Jones remains ignorant. Smith thus lacks practical freedom when it comes to voting and will give an unfree vote for the Evil Party.⁵⁸ If Smith knows for certain that it is absolutely impossible to resist

⁵⁸ Smith’s situation should not be confused with my imagined situation where I had a Pocket Oracle or became clairvoyant. The latter situation is about knowing in advance what I, based on my preferences, principles and so on will decide. The former is about knowing that something outside of my preferences and principles will force me to vote Evil against my will if I try to vote Good. Although I am uncertain how the former situation would affect my practical freedom, the latter situation certainly entails that my will lacks efficacy.

the device, we cannot pass moral judgement on his choice, because he has no choice; in this situation, he is simply not a moral agent. Even if he voted for Evil while the device remained dormant, he decided to do this *knowing* that he could not do otherwise, that his will was inefficacious. (If Smith had decided to vote Evil first and then learnt about the device, the choice would have been made under practical freedom and thus open to certain moral judgements.) Jones, however, retains some practical freedom as long as the device is not activated to compel him, since he *believes* that his will is efficacious when it comes to voting. Since Jones has some measure of practical freedom, he also retains some moral agency. How *much* of a moral agent he is with the device in his brain will depend on one's ethical theory. According to the virtue-ethical theory Jones might do wrong when voting Evil simply because he thereby manifests an evil character; according to subjective utilitarianism Jones might do wrong because, given his information, there was another option with better expected consequences. He is different from most moral agents in that a moral adviser could not tell him what he ought to do in this situation, but he might still have done wrong, and he might also be blameworthy for what he did. If we assume objective act utilitarianism, an action was wrong if and only if there was no better option that the agent could have performed if he tried to. In this case, Jones did *not* do wrong when voting Evil. But there could still be some kind of limited moral judgement to pass. For instance, if making this evil choice made Jones more inclined to choose the evil option in the future, the *choice* would have been morally wrong, even though the action was not. This is why the belief qualification is important. Having a will that is *truly* efficacious is more important for moral agency according to some ethical theories than others, but regardless of one's ethical theory it is implausible that a person in Jones' situation would be completely devoid of moral agency just because he has this dormant device in his head that he does not know about.

The reason why there is no belief qualification in front of “must choose” is that in the few situations where we can be mistaken about us having to choose, it seems obvious that we are not moral agents either. Suppose, for instance, that I am dreaming, and think I must choose between two options, but before I have time to make my choice the whole scenario is replaced by something completely different and the situation of choice evaporates. Or suppose poor Peter thinks he must choose (and, by implication, thinks he is able to choose) between taking the right path or the left one. However, in the very next second he is struck dead by lightning, so he did not have time to make a choice. It is quite obvious that there is nothing to pass moral judgement on here, since not only was there no action but also there was no choice. Barring a few special cases like this, however, we are not mistaken about whether we have to choose. We can be mistaken about the efficaciousness of our wills, but we know when we have to make up our minds about something. There is a “must” choose rather than merely “can” choose because there is really no getting around choosing our actions once we have reached the level of self-awareness where we can question our desires. I could try to refuse to make a choice by just sitting there, doing nothing, but that would be a choice already. I could decide to just throw a die in every situation of choice, but choosing to throw the die is a choice already, choosing to obey it is a second one.

In summary, minimal practical freedom is something agents have as long as they *believe* their wills to be efficacious, and must choose what to do. This is sufficient, freedom-wise, for having some kind of moral agency, although it may be limited – for instance, there may be things agents subjectively ought to do but nothing a moral adviser could tell them (as Jones in the earlier example), or it might be the case that only their choices but not their subsequent actions were wrong (as might be the case with Jones if we suppose act utilitarianism to be true). This concludes what I have to say about *minimal* practical freedom. Let us move on to the maximal conditions.

MAXIMAL PRACTICAL FREEDOM: Agents have maximal practical freedom if and only if they have minimal freedom, their wills really *are* efficacious, they have full information about their considered options, and all their unconsidered options are such that their own values do not give them an all-things-considered reason to pick one of these options over the one they actually chose.

An agent whose will really is efficacious has more practical freedom than an agent who merely believes so. If Jones and Smith decides whether to vote Good or Evil without having devices installed in their brains, they are open to the full range of moral judgements, appraisals, criticisms, blames, advice and demands. It is not just their choices, but also their actions, which can be judged good or bad, praiseworthy or blameworthy, and not just according to some normative-ethical theories but according to all. The information clause serves a similar function. Suppose that Jones and Smith decide whether to vote for Party A or Party B, and there are no devices in their brains, but all they know about the parties are their names. They still have minimal practical freedom, since they are in a situation where they have to make up their minds and choose. It is still not much in the way of freedom, however, to pick an option at random since one lacks information on which to base one's decision. Suppose, instead, that they do have information about the parties, but false information, leading them to choose party A while it is actually party B:s program that coheres with their values. Since they still have minimal practical freedom, they still have some minimal moral agency. According to some ethical theories there can still be some kind of moral judgement to pass (according to objective consequentialism, for instance, they can still do right or wrong). However, once Jones and Smith gain complete information about respective party, they will have full moral agency, full moral responsibility and be open to the full range of moral judgements of all ethical

theories. One might of course question what “complete information” means. Plausibly it does not include pointless facts such as the eye colour of the party members. That Smith and Jones have complete information about the parties should rather be interpreted as them possessing all information that *they* consider relevant. When this condition is in place, lack of information is no obstacle to their choice. If they also have information that they consider irrelevant, this will make no difference to their deliberation and choice, and it can therefore not enhance their practical freedom.

When agents have maximal practical freedom to choose between A, B and C, it is not merely the case that they “can” perform all of these options on a mere conditional analysis of “can”. It is not merely the case that they could do each of these options if they chose to. That agents can do, for instance, A, if they would choose to do A, is consistent with them having some kind of psychological barrier against considering A. The condition about unconsidered options in the above definition guarantee that the agent with maximal practical freedom “can” do his various options in a stronger sense than the merely conditional one, although the definition never leaves the practical perspective.

3.2.2. Unconsidered Options

“Unconsidered options” are alternatives where the efficacy condition is fulfilled, but which the agent either does not think about at all or think about in fantasy rather than in deliberation. The efficacy condition might be fulfilled; it might be the case that I believe that I could perform the action if I tried to, even if I do not actually think about the action in question. For instance, even if I do not consider watching a movie on Sunday afternoon, I might have beliefs regarding my economy, the price of a movie ticket, the location of the theatre, the commuter train and so on that, taken together, imply that I could see a movie Sunday afternoon if I chose to. In this instance I think it is plausible to count the efficacy condition as fulfilled and say that I do believe

that I could see a movie Sunday afternoon if I decided to do so, although I have not thought about the matter at all. When it comes to the difference between fantasising and deliberating further clarification is needed. Mathew Noa Smith has defined the difference between deliberation and mere first person fantasies as follows: Deliberation ends with an intention, or at least that is what we aim at when engaging in deliberation.⁵⁹ A first person fantasy does not. When first-person-fantasising about performing a certain action, my thoughts may be just as vivid and detailed as when deliberating – the difference is merely that I do not aim at arriving at an intention.⁶⁰ Admittedly there is probably a grey area here, where daydreams gradually slide over into deliberation, but as long as there are clear cases of first-person fantasies that are *not* deliberation the distinction remains important. Smith's thesis is that an action must fit into our personal narrative through which we construct our identity to be an object of deliberation. However, I will not discuss what kind of psychological mechanisms there are that might restrict a potential action to the realm of first-person fantasies. What I *will* discuss is when this restriction is freedom-undermining. Practical freedom is the freedom to choose one's actions, and this freedom is limited when one's ability to choose is compromised. I will argue that agents aim to perform the action, out of the set allowed by the efficacy condition, which is best according to their values. When engaging in deliberation, agents try to figure out which option is the best one, and choose it. Therefore, agents' ability to choose is compromised when they *fail* to consider the option that is best according to their values. If they do not consider an option, they cannot choose that option either, and thus cannot do what they attempt to do when engaging in deliberation. When this happens, their practical freedom is limited.

The claim that agents always try to pick the action that is best according to their values may sound strong and controversial, but it is not intended to

⁵⁹ Smith, 2010, p 7,

⁶⁰ *ibid* pp 6-7

be.⁶¹ By “value” I mean to include anything that they consider reason-giving. It could be broad goals in life such as achieving greatness within one’s career or loving and being loved by one’s family. It could be something much more local and small; for instance, somebody may value having a good time on Saturday night. That person would then think he or she had reason to do something that he or she believes would promote having a good time on Saturday night. It might be that values in this very broad sense must always be based on desires or preferences, that what I value is always something I also desire (the alternative view is that they could, at least sometimes, be purely intellectual and unemotional), but even if that is true, it is not the case that all desires or preferences give rise to values. I can have a desire without regarding it as giving me any reason at all to do anything. I may consider the desire in question unwanted, something one should try as far as possible not to pay attention to. This is why deliberation and choice is different from mere introspection; just finding out which desires I have can be achieved by introspection alone, but taking some desires as reason for action is something active, something I do. As Wolf puts it “We can [...] construe the notion of values somewhat more broadly, as comprising those things which a person cares about, or alternatively, as including all and only those things which *matter* to a person in some positive way, without losing the distinction between what a person values and what she merely wants, or desires, or likes.”⁶² I do not have to assume that one of my options is good in a very substantive sense of the word in order to make a choice, but I must in some loose sense value (and not just desire) some things over others. Some choices are made rather indifferently, since there is no particular reason to take

⁶¹ Some philosophers, like David Enoch, would think the claim too weak. He thinks that all deliberation aims at is finding out which action, in a substantive realist sense, is the best thing to do in this situation. Personally, I do not agree, but his claim is consistent with mine. Agents who aim at finding the right thing to do in a robust realist sense will attempt to make their values conform to what they take to be normative facts, but then they will still choose the action that is best given what their values, which is to conform to the normative facts.

⁶² Wolf, 1990, p 31

one option rather than the other, and yet something has to be picked, but even these “indifferent” choices cannot be made unless one values making a choice and be done with it over just standing there like Buridan’s ass. Other choices are such that one alternative is obviously better than the others, according to the agent’s values. Then again some choices may be very hard, and making up one’s mind in these instances may be described as simultaneously choosing how to act and how to value; one chooses to consider some reasons as stronger than others and thereby chooses how to act.

When “value” is construed in this way, it becomes almost trivial that we always try to pick the option that is best according to our own values. It is not *completely* trivial, since it is at least possible to imagine an agent to whom “having fun” matters in a positive way, to whom *nothing else matters* in this particular situation, and yet consciously picks a boring option. In real life we might very well choose to do something that is boring or that we do not like to do, but not if *fun is all that matters*. It is an uncontroversial fact about real human beings that when we choose what to do, we do try to pick the option that is best according to our own values, when values are construed in the broad way suggested by Wolf, as anything that matters to a person in a positive way.

It might be objected that we do not aim at picking the option that is the very best one, but only one that is good enough. I do think we aim at picking the best one if one factors in the cost of continued deliberation in the value of the option. Suppose I am going to buy a German Shepherd puppy, and choose between different dog breeders. I investigate various breeders regarding the health and mentality of their dogs, how well they care for their dogs, how much of an interest they take in the welfare of already sold puppies, etc. Now if I were to investigate all breeders of German shepherd dogs in Sweden to find the one who scores highest under all criteria I consider relevant, I would probably have to devote so much time to the search that my job would suffer, and I value doing well at work. Therefore, choosing the breeder who

scores highest according to all relevant criteria is not the best option according to my own values. The best option is actually to devote some but not all of my time to the search, and then pick a breeder who scores highly, although not necessarily the highest, according to the criteria.

Now as Smith writes, if there are a number of actions that are possible in the sense that agents think their wills are efficacious regarding them, this does not mean they can actually think all of them through. In every situation of choice, there are millions and millions of things people could have done had they wanted to, many of them completely pointless – things like pulling their eyebrows together, sticking their tongues out, making various noises etc. There will also be actions they cannot seriously deliberate about simply because they are evil – like throwing a baby out of the window to a certain death. It would be absurd if a definition of freedom entailed that their freedom was constantly being diminished by the fact that they fail to consider pointless and evil actions before making a choice. The value-qualification about unconsidered options prevents this absurd result. Since agents who choose try to pick what is best according to their own values, it is no obstacle to their choices that millions of options are left unconsidered, as long as they were not better according to their own values than the one they picked. Thinking through countless of options is simply not necessary for finding out which one is best, or at least one of the better ones, according to one's own values. If Philip values a fun Saturday night, those things he finds most "fun" will probably strike him as obvious alternatives, while all those things that he thinks his will would be efficacious in realizing if he chose them but which he does not regard as "fun" are simply not thought of. Thus, even if he just consciously thinks of two or three options out of potentially millions that he could have performed had he decided to, he may have a really good chance of finding the option that is best according to his personal idea of fun. The same is probably true about Pereboom when he deliberates about his

future career, in his article “a Compatibilist Account of the Beliefs Required for Rational Deliberation”.

Pereboom argues, against Kapitan, that although the presumption of contingency is necessary for rational deliberation, it is not sufficient together with an efficacious will. A third condition is needed; besides not being able to deliberate about things that the agent *knows* will not happen, he cannot either deliberate about things he *supposes*, at least for the sake of deliberation, will not happen. As Pereboom puts it, “An agent regards a proposition as settled just in case she has no doubt that it is true, or else, she disregards any such doubt she has, e.g., for the purpose of deliberation.”⁶³ Pereboom’s own example of something that is settled for him is that he will not give up his philosophy career to go to Africa and become a mercenary – this is simply no option for him, because he takes it for granted that he will not do this.⁶⁴ Otherwise it fulfils Kapitan’s conditions for deliberating; it is contingent relative to his beliefs (nothing he knows really *entails* that he does not go to Africa to become a mercenary), and he also thinks that he could do this if he would decide to. However, it seems to me that the most plausible explanation of why Pereboom does not deliberate about becoming a mercenary, is not that he *believes* that he will not become one, but that he has no good *reasons* for giving up his philosophy career and going to Africa. (If Pereboom had good reasons to become a mercenary in Africa, and thought he could do this if he tried to, he would probably not be so quick to consider the proposition “Pereboom will not go to Africa to become a mercenary” true either.) Suppose Pereboom values such things as having an interesting, safe and well-paid job, and being a philosophy professor satisfies these to a high degree. It is then likely that this is the option that is best according to his values, or at least one of the best options. If there was another job that were, for instance, much more *interesting* according to Pereboom’s own idea of

⁶³ Pereboom, 2008, p 7

⁶⁴ Pereboom, 2008 p 13

interesting, he would likely have thought about it. I am not discussing counterfactual values here. There may be another Pereboom in another possible world who, through a series of events that never happened in our world, became completely immersed in the world of plants, went on to become a botanist, and who finds botanics much *more* interesting than the real Pereboom finds philosophy. However, the issue is not whether Pereboom may have preferred some other job if he had had different values, as if his idea of what is “interesting” had been different from what it actually is. When Pereboom deliberates about his future career he tries to make up his mind regarding which job would be best according to *his* values, not those of some counterpart in another possible world. The fact that there are countless jobs he never even considers is not necessarily a problem for that. He may still choose his future career under maximal, or close to maximal, practical freedom.

Practical freedom is thus only diminished by unconsidered options when at least one of the options is left unconsidered despite being *superior* according to one’s own values to the option one chooses. Here is an example: Suppose that Wilma, a girl from a working class background, deliberates about her future. Wilma’s parents are blue-collar workers, and so is everybody she knows. She does not have one family member or friend who went to university. Wilma sometimes fantasises about going to university, it seems interesting and fun to her, and her dream job is something one needs a university education for. However, the only jobs she deliberates about are typical blue-collar jobs. She eventually follows in her parents’ footsteps, despite not liking the job very much. Let us suppose that Wilma did have grades good enough to apply for university. If she did not have that, there is an obvious and non-controversial sense in which she is not free to do that. Let us suppose that there were no material obstacles preventing her – that would be true enough if she lived in Sweden, where university education is free, and

where anyone can take out a loan to support her while she studies. Many of us probably still have the intuition that the choice to follow in her parents' footsteps was not quite a free one, although it is difficult to say exactly where the lack of freedom lies. One may talk about class and structures and social groups, without really defining what makes this particular individual unfree. Wilma falls short of maximal practical freedom because there was at least one option regarding her future, namely "going to university", which was left unconsidered despite being superior (according to her own values) to the choice she actually made. Betty, from a similar background, does not seriously consider going to university either. However, Betty does not long for an academic job and is perfectly happy with the job she actually chooses. Therefore, her freedom is not diminished. Similarly Allison, who values going to university and also does that despite her working-class background, is fully free as well. What factors caused Wilma, Betty and Allison to become the way they are is outside the scope of this dissertation. It is a matter for empirical sciences like psychology and sociology to decide. I am merely concerned with distinguishing between people with more or less free will – not with explaining how they got there.

A more dramatic example of people with diminished practical freedom is given by Stanley Milgram's famous obedience experiments. The test leader told his subjects that they were participating in a study on learning and punishment. One subject was to be strapped into an electric chair while the other subject was to control the shocks. If the first subject misremembered a list of words, the second subject would shock him on the test leader's orders. Actually, the subject in the electric chair was Milgram's accomplice and did not really receive any shocks. Most of the real test subjects, the ones who were to administer "electric shocks" to an innocent victim, obeyed the test leader,

even to the point where it seemed like they were seriously endangering the victim. Most of them felt terrible about what they did, and yet they obeyed.⁶⁵

I think the difference between the obeying majority and the disobeying minority can be accounted for in terms of practical freedom. Now there is one obvious way in which *all* of the Milgram test subjects lacked maximal practical freedom; all of them lacked crucial information about what was actually going on. None of the test subjects really knew what one's options were. Let us, however, for the sake of discussion, disregard the fact that the test subjects were fooled. Imagine that the victim really was electro-shocked, just as the test subjects thought. Now most people do think that electro-shocking innocents just because an authority figure tells them to is morally wrong. For most people it is thus the case that disobeying the test leader and refusing to shock the victim would have been the superior alternative according to their own values. Then why did most people not disobey? I think the most plausible answer is that they did not *seriously think* of this option. This is also Milgram's own hypothesis: The people who obeyed the test leader simply failed to ask themselves the seemingly obvious question: Why should I serve the person who came up with this test and let the victim suffer?⁶⁶ The small number of individuals who did stand up against the test leader and refused to participate might not have been *kinder* or more *empathic* than the others. They were just the ones who managed to see clearly that they had two options: Either disobeying this cruel and mad scientist and do the morally right thing, or obeying the cruel and mad scientist and electro-shock an innocent person. As soon as one realises that these are the two alternatives, it is quite obvious that the first alternative is the best one, according to almost anybody's system of values. However, to most people it might have been the case that the first option was left unconsidered, simply because they could not take seriously the idea of flat-out disobeying an authority figure like

⁶⁵ Milgram, 1974, pp 25-38

⁶⁶ Milgram, 1974, p 21

Milgram's test leader. According to PPC, the difference between those who obeyed and those who did not would be a difference in free will. A minority managed to make a free choice in circumstances where most people's freedom was severely diminished. Moreover, if their freedom was diminished, so was their moral agency. What it means that their moral agency was diminished will depend on one's normative-ethical theory. If we assume a theory that uses the concept of moral responsibility, we could argue that the test subjects who obeyed the test leader had diminished (or perhaps no) moral responsibility for what they did, and were not fully blameworthy. One could also argue that in the morally relevant sense of "can" people cannot do what they cannot deliberate about, and since ought implies can, people do not have a duty to perform what is an unconsidered option to them. (I am not saying this is how one ought to think about the ought implies can principle, but it is a possibility worthy of discussion.) If that were the case, a utilitarian should only consider options such as trying to help the victim remember the right word, pleading with the test leader and so on, options that do not include straight-out disobedience, when calculating the consequences of different alternatives for the test subjects for whom disobedience was left unconsidered, and a deontologist should not consider disobedience a duty for these test-subjects.

On the other hand, those test-subjects who managed to make free choices in this difficult situation had full moral agency. They were fully morally responsible, fully praiseworthy, and managed to perform their duty.

3.2.3. More or Less Practical Freedom

Maximal practical freedom is something an agent might have in particular situations of choice, although nobody is perfectly practically free throughout his entire life. However, if we hold up maximal practical freedom in every choice made throughout an entire life as an ideal, then some individuals will come closer to this ideal than others. I will not go into the debate on what

constitutes rationality, but an individual who always chooses under maximal practical freedom would probably count as highly rational according to most theories. Individuals whom most people would regard as irrational, who has some very strange way to deliberate, could still have maximal practical freedom in a single choice. They could still, by pure luck, manage to choose the alternative that is best according to their own values. If so, they were (also by pure luck) perfectly free in that choice, in that particular instance their irrationality was no obstacle to them. Even so, an individual who is always free, who reliably chooses the best alternative, would have to be somebody who deliberated in a rational manner.⁶⁷

Agents who have some false but mostly true beliefs regarding the efficacy of their will, will fall somewhere between minimal and maximal practical freedom, and the same goes for agents who have some false but mostly true beliefs regarding their considered options. Agents who deliberate over some options, but who do not think of or merely fantasizes about one or a few options that would have been better according to their own values than the one they eventually settle for, also fall somewhere between minimal and maximal practical freedom. I do not have an exact formula for comparing different individuals when it comes to this latter dimension of freedom. Perhaps it is the number of unconsidered options better than the one the agents choose that determines their level of freedom; the higher the number, the lesser the freedom. For instance, if I chose to study philosophy but failed to consider going to art school, although art school would have been better

⁶⁷ It seems conceivable that an individual who deliberates in a strange and irrational way happens to choose, by pure luck, the best alternative *every time* he makes a choice throughout his entire life. I do find it counter-intuitive to claim that this confused but lucky individual is a full-blown moral agent. I am prepared to accept this claim though, since I think the feeling of counter-intuitiveness may be due to it being difficult to imagine this scenario. We have stipulated that this agent is so lucky that he always does what is best according to his own values, values that must be perfectly coherent since he is just as lucky when choosing his values as his actions (I write about free choice of value as coherent choice of value in 3.2.4.). Does it even make sense to label such a coherent individual irrational or confused? Does it really make sense to say that he constantly picks the right action by pure luck, just because he deliberates in an unusual manner? I am uncertain.

according to my own values, my freedom is a little bit diminished, while if I chose philosophy and failed to consider either art school, becoming a nurse or a professional dog trainer, and all three options would have been better according to my own values, then my freedom would be more diminished. Perhaps it is how much better the unconsidered options are than the one actually chosen, and if they are just a little bit better, then freedom is not diminished as much as if they were very much better. For instance, if I went to see a movie Saturday night and never thought about going to a certain club, although the latter would have been slightly better according to my own values, my freedom is slightly diminished, while if I electro-shocked an innocent person and never thought about disobeying Milgram's test leader, although the latter would have been much better according to my own values, my freedom is severely diminished. Perhaps it plays a role how distinct the unconsidered options are from each other and/or from the option that gets chosen. For instance, if there are five movies I failed to consider despite all of them being more to my taste than the movie I actually go and see, my freedom is still not very diminished, since the options are all various movies and therefore not very different, but if I failed to consider going on a hike instead, and that would have been better according to my own values, my freedom is more diminished. It could be a combination of all of these factors, but to come up with a precise formula lies outside the scope of this dissertation. In any case I want to leave room for the possibility that an agent must come fairly close to the ideal of a perfectly free agent to count as a full-blown moral agent.

3.2.3. Diminished Practical Freedom Compared to Weakness of Will

Some readers may wonder how diminished practical freedom due to unconsidered options relates to weakness of will. The difference between suffering

from weakness of will on the one hand, and a diminished practical freedom on the other hand, is that weakness of will occurs late in deliberation, while diminished practical freedom means that some options will not enter deliberation in the first place. Weakness of will as traditionally understood means that I judge X to be the best alternative, and I then succumb to temptation and fail to do X anyway. For instance, I judge that writing three more pages on my dissertation is, all things considered, the best thing to do right now, but I end up spending a couple of hours corresponding with my friends on Facebook. I simply give in to temptation and have my sound judgement overridden by laziness. Nevertheless, writing three more pages was a real option for me, something I deliberated about. It was no idle first-person-fantasy like going to university was for Wilma in an earlier example.

Philosophers disagree on whether weakness of will is a mitigating factor in wrongdoing or not. I think the idea that wrongdoing is less blameworthy and rightdoing more praiseworthy if it was *difficult* to do the right thing is plausible, and can be incorporated into most theories of free will and morality, including PPC.⁶⁸ If so, weakness of will could be mitigating, depending on the exact circumstances of the case. If neurosis, like the often-cited example of a kleptomaniac who steals despite judging that it is better not to, is counted as a kind of weakness of will, this is plausibly a case where doing what you judge best would have been very, very difficult. If so, then the kleptomaniac is only a bit blameworthy for stealing, and would be very praiseworthy for resisting his or her urge. The same line of reasoning can be applied to addicts. The limiting case of difficulty is of course impossibility; perhaps there are some people so neurotic that doing something against their neurotic urge is psychologically *impossible*, not just very difficult, and if so, it is not the case that they ought to either. On the other hand, when I spend several hours on Facebook out of sheer laziness, I would be as blameworthy

⁶⁸ See also McKenna's discussion of this idea in McKenna and Russell, ed, 2008, p 215

as anybody who spends time on Facebook instead of working, since it would not require that much teeth-clenching to resist.

The idea that it is impossible to do otherwise but fall for temptation only in the case of *extreme* neurotics or addicts may seem to beg the question against incompatibilism. But it does not, since I use impossibility as a limiting case of difficulty. Difficulty, in turn, should be understood in terms of general abilities, as a measurement on how much effort an action requires when we do undertake that action. For instance, I have the general ability to lift five kilograms above my head, since if I try to, barring special circumstances, I will succeed. It is also something that is *easy* to do, since I do not have to tense my muscles very much to do so. I also have the general ability to lift twentyfive kilograms that way, but that is significantly more difficult, since it requires tensing my muscles substantially. I lack the general ability to lift fifty kilograms that way; no matter how hard I tense my muscles, I do not succeed. Between twentyfive and fifty there will be weights that are very difficult for me to lift, but not impossible, weights I might succeed in lifting by pushing my muscles to the limit, but not otherwise. Now think of not stealing in analogue with weight-lifting. For a regular person, not stealing may in most situations be like lifting a five kilogram weight or perhaps nothing at all, while for the kleptomaniac, it might be like lifting a twenty-five or thirty kilogram weight. For some kleptomaniacs it might be analogous to the fifty kilogram weight. Just as I simply do not have the arm muscles to lift that much, so too the kleptomaniac may simply not have the mental muscles to resist an urge as metaphorically heavy as this one. In that case, stealing will not be wrong; there will be no other alternatives available to the kleptomaniac with which we can compare the consequences, the kleptomaniac can have no duties to resist stealing and she cannot be morally responsible for doing so (at least assuming that she cannot get around the problem by avoiding to be near anything stealable in the first place). According to some ethi-

cal theories it might be the case that she still ought to *try* to resist, despite it being fruitless, but her failing will not count as morally wrong.

Thus far I have assumed that weak-willed persons *know* their own weaknesses, or at least have a fairly good grasp of them. It might of course be the case that neurotics or addicts are deluded and think they could easily give up stealing, drinking or whatever their neurosis or addiction is about, while it would in fact require such a great mental effort so as to be impossible for them. From their practical perspective, they choose between quitting and continuing as before. Let us suppose that quitting is no unconsidered option, but one they seriously think about (granted, this may not be the case for most neurotics or addicts, for many of them quitting may be an unconsidered option). If so, they are at least minimally practically free to choose between quitting and continuing; they face two options and must make up their mind about them. If they choose to quit, they will of course discover that it was impossible after all, and therefore this is one of those cases where practical freedom seems not worth wanting, since it rests on an illusion about the efficacy of one's will. We can still pass a kind of limited moral judgement on their choice; we might argue that it ultimately has better consequences if they choose to quit (if they make this choice they will learn that their addiction is after all irresistible, and then they can deliberate about whether to go to a rehabilitation clinic or ask friends and family for help), or that it was right because it reveals a good will on their part, and we can praise them for at least trying.

Thus, weakness of will, defined as difficulty or inability to carry out one's decisions, only completely destroys agents' practical freedom when the weakness is so great that it makes it impossible to carry out certain decisions, that is, where it makes their wills inefficacious, and the agents know about this. If agents falsely believes that their wills are efficacious they are still free to choose, although freedom in this situation may seem a freedom not worth wanting, and the kind of moral judgement we can pass will be

limited. If the weakness merely makes some actions that are easy for other people difficult for them, then the weakness changes the nature of the options they choose from. Instead of choosing between, for example, easily abstaining from theft and succumbing to temptation, the agents choose between abstaining from theft through an enormous effort of will and succumbing to temptation. This affects how praiseworthy/blameworthy they would be for taking a certain option, but there is still choice, and still practical freedom.

To contrast weakness of will with diminished practical freedom I will once again make use of Milgram's obedience experiments for illustration. I wrote earlier that a plausible explanation of what happened is that the obedient test subjects had their practical freedom diminished, while the disobedient minority clearly saw that disobedience was an option. Suppose, however, that there were some people who did see disobedience as an option, and judged that it was clearly best to disobey. Then the test leader gives his next order, and the subjects become so intimidated by him that they become unable to voice their opinions; their cowardice is simply stronger than their sound judgement. That would be weakness of will rather than diminished practical freedom.

In practice there are situations where these phenomena are difficult to tell apart, but when thinking through one's own experiences it is usually not difficult to know what is what. There is a clear difference between on the one hand making a judgement as to what is best but failing to act on it, and on the other hand just going along with something because one never seriously considered any other option. Many everyday instances, like getting stuck on Facebook instead of working, are clear weakness of will cases – if one has a job, the option to work is rather obvious. Diminished practical freedom is also easy to diagnose in those situations where it is either obvious from the way people talk about their plans that they are engaging in first person fantasies rather than actual deliberation, and in situations where peo-

ple go against their own sincere values despite neither threat nor temptation being present. Some cases, however, are difficult to diagnose, and some cases may be borderline.

I have shown how the theory of unconsidered options implies that agents have a diminished practical freedom in many situations where it does seem plausible that people's moral agency is somewhat compromised, for instance with Milgram-type situations and with people who have been shaped by their background to regard certain choices as obvious. Neurotics and addicts, two of the textbook examples of people whose status as moral agents is in doubt, may in some instances fail to satisfy the efficacy condition and thus lack practical freedom. Even so, it is probably much more common that they do have practical freedom; their options only look slightly different from the options of non-neurotics and non-addicts. "Not stealing", for instance, is a simple matter for most people, but may be something that requires a huge mental effort for a kleptomaniac. To delve deeply into the problems of addiction, neurosis, and so on is outside the scope of this dissertation; I am just pointing roughly in the direction a solution would lie, according to PPC.

Let us now turn to psychotics, the final text-book example of people who cannot be responsible for what they do. Psychosis can take many forms, and not all would have the same moral-philosophical implications. A psychotic person who experienced himself as a helpless passenger in a body that acts driven by something other than his will, has simply lost his practical freedom completely and is no moral agent at all in my view. A person who does act and experiences himself as acting, but who has a wildly distorted view of the world, is a different case. In some cases, it would make most sense to regard him as a moral agent, but one who lacks a significant amount of crucial information and (non-culpably) has many false beliefs. A non-psychotic person who, for non-culpable reasons, has wildly misleading beliefs and then decides to act on those will, of course, be judged differently from a person

who performed the same actions with true beliefs. For instance, if I, for non-culpable reasons, have the false belief that stealing my neighbour's car is the only way to save a child in mortal danger, I should be judged differently from somebody who steals his neighbour's car fully knowing that refraining from the theft would not have resulted in anyone dying. It might be that I did what was subjectively yet not objectively right and that I should not be blamed for what I did. Thus, having a non-culpable false belief is morally relevant, but it is not necessarily morally relevant whether I had the false belief because I was subject to elaborate deception by other people or because I had psychotic delusions.

All in all, psychosis is a broad concept that covers many kinds of symptoms. Even if it is true in general that psychotic people cannot be subject to the same obligations as healthy people, the precise reasons for that may vary from case to case, and it is not necessarily the case either that a psychotic person completely lacks moral agency. These questions would have to be settled on a case by case basis, depending on the exact symptoms of each individual.

3.2.4. Free Choice of Values

We do not just choose actions. We sometimes choose what to value and principles to act on as well. An agent who deliberates properly about some options in light of values that have never occurred to him to question would lack something freedom-wise. An ideal agent with perfect practical freedom would not just choose all his actions, but also all his values, under maximum practical freedom.⁶⁹

⁶⁹ An agent with perfect practical freedom might seem impossible even in theory. It is hard to see how one's very first values could have been *chosen*. One would need to have some values already in place in order to choose new ones. But suppose an agent has some values installed in him by his parents, choose to value further things based on those first values, and later reflects over and choose to keep his first values. Such an agent could have maximal practical freedom.

As in the case of actions, we cannot deliberate about all things that fulfil the efficacy condition. But how could a value fulfil this in the first place? My will would be efficacious in adopting a new value if I could adopt this value if I chose to. I will not offer a precise definition of what it means to have adopted a new value, but a definition should include acquiring new behavioural tendencies and/or certain deeply felt emotions. If I cannot adopt value V at will, since even if I tried to I would not be able to modify my emotional life and/or let my behaviour be guided by V, this means that my will is not efficacious when it comes to the adoption of V. If I know this, V will be outside the scope of deliberation. It may also be the case that V is an unconsidered option. An action left unconsidered is freedom-undermining if the action was better according to the agent's own values than the one he or she ended up choosing. In a similar way, it is freedom-undermining when a value is left unconsidered although the adoption of the value would lead to greater coherence in one's total system of values and value-related beliefs. A person with maximal practical freedom is thus a person with as much coherence as possible among his or her values and related beliefs, who chooses to perform those actions that are best according to said values. If a person on the other hand is unable to acquire value V either because his or her will is not efficacious on that point or because V is left unconsidered despite cohering with his or her other values, we might say that it is not the case either that he or she ought to acquire V, at least not if we use ought in a subjective sense, and that he or she is not morally responsible or has a mitigated responsibility for not acquiring V.

Suppose, for instance, that Simon grew up in a thoroughly sexist society. He actually believes in the value of many traditional virtues, like benevolence and showing respect for other people, which cohere poorly with his society's patriarchal traditions where men do not allow women to decide for themselves. He has had some brief encounters with non-sexists, that left him completely shocked by the fact that there are people with such strange views,

and unable to seriously listen to their arguments. Simon may entertain some brief first-person-fantasy of living in a society with equality between the sexes after such an encounter, but he does not seriously deliberate about it. It is just an impossible fantasy scenario to him, and he remains a convinced sexist. Sexism is therefore not something Simon has chosen. Simon would then have diminished moral responsibility for his sexism, the consequences of him becoming a feminist might be seen as morally irrelevant since he cannot deliberate about becoming one, or he might have no duty to become a feminist (but perhaps still some weaker duty to try to modify patriarchal patterns of behaviour in the right direction, if this is within the scope of things he can deliberate about). Now compare Simon to Eric, who also lives in a sexist society. Eric is convinced that might is right, and the mere fact that one group in society manages to oppress the rest is enough to justify the oppression. The reason that Eric does not listen to people who argue for the equality of the sexes is that he does not agree with their premises of benevolence and respect. He can see that becoming a feminist and eventually recasting society in a more feminist mould is a real option, he just does not agree with it. Now it might be the case that Erik is unfree on some deeper level; perhaps his might-is-right-view itself was not chosen under practical freedom. However, assuming for the sake of argument that it was, and assuming for the sake of argument as well that Eric does have whatever other abilities are required of a moral agent, he ought not to be a sexist, and is fully responsible for being one. Simon has failed to *choose* his values when it comes to women, while Eric is a freely chosen sexist.

3.2.5. The Things We Just Do

That individuals who have practical freedom *in general* may fail to have this in a particular situation is obvious, if we talk about such things as being carried away by a strong wind. But what about all the situations where we sort of just do things? This could pose a problem for theories that stress *choice* in

the way PPC does. It is true that we often choose our actions, and that it is necessary for us to make countless choices each day, but far from all intentional actions are preceded by a choice. Wolf also mentions the fact that we often just do things as a potential problem for choice-centred theories: It is not like one really pours a second cup of coffee or put on a sweater based on some special values one holds.⁷⁰ There seems to be something deficient about a theory that claims that actions can be neither right nor wrong despite being intentional, simply because we didn't *choose* them, but sort of just did them anyway. (When it comes to our values, it may be even more obvious – often we just value some things, without exactly having chosen to do so.)

Let us use as an example how I walk from the underground station to the university. This short walk is not something I actively choose to do. I do not consider, even briefly, to take another route instead, or go to a different place. I do not make a conscious choice to take the route I take. I just go there.

I think there are several possible ways to handle these actions. Sometimes the action could be part of a larger action that was chosen after all. My short walk between the underground and the university building could be part of the larger action “going to the university”, and I might have thought about, say, whether to go there, work at home, or take a day off and work Saturday instead. Sometimes the action could have been chosen once upon a time, and then slid into a habit I no longer think about. For example, the first time I went to the university I might have considered different routes between the underground station and the university buildings, and picked the one I deemed best, but eventually it just became habit. In a way though, the route I take is chosen by me, since I chose it once upon a time. However, I do not think that all intentional actions that we sort of just do can be incorporated into one or both of these categories.

⁷⁰ Wolf, 1990, pp 32-33

I think the right way to deal with all actions that we just do, whether they also are once-chosen habits and/or part of larger chosen actions or not, is to say that they are, insofar as they are intentional, the result of a kind of “choice”⁷¹ where all alternatives except this one is left unconsidered. How free such a “choice” is depends, as in all choices, on whether some of the unconsidered alternatives would in fact have been better according to my own values or not.

This would be true for values that we ostensibly just have as well. Suppose that Freddy, unlike Simon and Eric, grew up in a society where there is full equality between the sexes. Suppose that Freddy, just like Simon, values benevolence and respect, but unlike Simon he also thinks women should have the same rights as men. Freddy has never deliberated about sexism, since equality between the sexes is so obvious to him. Yet his belief in equality between the sexes counts as a freely chosen value, because it coheres with his other values. Sexism is an unconsidered value option, but not one that would have been better according to his other values, and therefore it is not freedom-undermining that he has not considered it. We might not praise Freddy for his belief in equality simply because that belief came so easy for him, and one only deserves praise for doing what is morally right when some kind of effort is involved. Still, he is morally responsible for so believing, and since he is a moral agent his belief can be morally right.

3.2.6. The Importance of Being Able to Act on One’s Own Values

Now somebody might object that it is strange that it is the agent’s capacity to act on his *own* values that determines if he can be a moral agent or not.

⁷¹ The quotation marks are there to signify that “choice” is now used as a term of art. In everyday language we only use this word about mental events where we actually thought of several options before picking one. I now use the term about mental events where we only considered the option we picked as well.

Should it not rather be the agent's capacity to act on *moral* values? Imagine a completely coherent egoist with maximal practical freedom when it comes to choosing actions as well as adopting values, who just cannot see the point of caring about anybody but himself. If he cannot even begin to grasp why it could be bad to steal even when you are not caught, or why helping others could be a good thing, is it really plausible to consider him a moral agent?

Now some philosophers would deny that a perfectly coherent egoist is possible. Perhaps some kind of contradiction is involved in thinking that I ought to maximise my own well-being, while a qualitatively and quantitatively similar well-being in another creature gives me no reason for action. Perhaps egoists simply have too many loose ends in their belief system to count as fully coherent; while believers in an altruist morality may explain why their own well-being matters (by pointing to factors that also underpin the importance of other creatures' well-being), egoists cannot explain why they are supposedly more important than others, but must simply take that as a given. Despite the fact that nobody has managed to produce a truly convincing argument to the effect that egoists cannot be fully coherent, I personally think there is some truth here. Even if perfectly coherent egoists are possible in theory, I think that as an empirical fact it would be easier to find a coherent altruist than a coherent egoist. However, let us put that to one side, and suppose, for the sake of argument, that egoists with full practical freedom according to my definition is possible. Now why does it not count as freedom-undermining that they are unable to appreciate moral reasons for action and therefore the morally right options in a situation? Additionally, can they really be moral agents?

It is important to remember when considering the second question that I do not claim that practical freedom is sufficient overall for moral agency. I merely claim that it is necessary, and that it is sufficient *freedom-wise*. Plausibly people need some capacities besides freedom in order to be moral agents, like the ability to relate to others in certain ways, to see that other

creatures can provide me with reasons for action even when they do not directly affect my own well-being. Our hopeless egoist might thus be a non-moral agent in virtue of lacking some non-freedom-related capacity. I still need to explain why an inability to appreciate moral reasons for action does not count as freedom-undermining, or why the ability to choose what is best according to one's *own* values should be crucial for moral agency. Freedom or free will is, roughly speaking, the ability to be the origin of one's actions and/or to have several options to choose from, although different freedom theories analyse these conditions in different ways. I argue that agents are the origin of their actions in the sense that they must choose them, they cannot delegate the choice to something outside them, and they will often choose between several options. An inability to appreciate moral reasons for action does not make one any less the origin of one's actions, so if this inability is freedom-undermining, it has to be because it limits the number of options one can choose from. Even so, as I have already pointed out, all people are severely limited in the number of options from which they can choose. In every situation there are millions of things I could do if I tried to, and yet sheer lack of time means that I cannot consider more than a tiny portion of them. If limitations on the number of options one can consider in deliberation in itself were freedom-undermining, nobody would have much freedom. It is only freedom-undermining not to consider those options that would have been best according to one's own values, because when this happens the agent does not manage to fully *choose*; he or she does not manage to do what he or she tries to do when engaging in a choice. This is morally relevant since morality is action-guiding; when agents fail to choose morality cannot guide their actions. If agents fail to pick the option that is best according to their own values, they will fail to do the right thing even if they have correct moral beliefs and values being moral. The fully coherent egoists on the other hand, do not face any obstacles when they are trying to choose. They set out to pick the option that is best according to themselves,

and this they manage to do. The fact that they cannot see the point of morality is thus no more of an obstacle for their choices than the fact that I cannot see the point of baseball is to mine. Therefore their *freedom* is not diminished, although psychological factors that lie behind their inability to grasp moral ideas may still be morally relevant, and might make them non-moral agents.⁷²

Just like the fully coherent egoists can have practical freedom, so too can perfect saints, who never really deliberates about what to do but always does what is morally right. If the saints are whole-heartedly good and always does the right thing without giving the matter much thought, because what the right thing is and that he should do it is obvious to them, it follows that among all the unconsidered options there is none that would have been better according to their own values than the actions they actually perform, and they can thus have full practical freedom.

I have now given an argument for the sufficiency of practical freedom for moral agency, and explained the conditions that need to be fulfilled for an agent to have practical freedom at all as well as maximal practical freedom. Perhaps the reader still asks herself what is so special about PPC, and how it is different from other compatibilist theories in the debate. The difference lies in PPC:s consistent adherence to a practical perspective, and the consequential disregard for causes of behaviour.

⁷² A non-moral agent cannot do right or wrong, and cannot be morally responsible even if compatibilism is true and the best moral theory employs this concept. However, I will not give a full definition of a non-moral agent, since that would require going into questions of what it takes to be an *agent*. We do not need to know exactly what an agent is in order to define what a moral agent is, since those who can do right and wrong count as agents on any theory. But is, for instance, a dog a non-moral agent? Or a simple robot, that can act somewhat independently but does not possess an impressive artificial intelligence? These are interesting questions, but they fall outside the scope of this dissertation.

3.3. Practical Perspective Compatibilism Compared to Causal Compatibilism

Most compatibilist theories are what I call “causal theories”. They view people from a theoretical perspective, ask what causes their actions, and then declare that some causes are compatible with moral responsibility, while others are not.⁷³ For instance, according to Fischer and Ravizza, determinism is no threat to moral responsibility. It is, however, necessary that one’s behaviour ensues from the right kind of “mechanism”.⁷⁴ The mechanisms that leads to behaviour is sometimes said to be physical,⁷⁵ sometimes it seems to be mental.⁷⁶ It is also important that the mechanism has the right causal history if it is to ground moral responsibility.⁷⁷ According to Alfred Mele, one’s behaviour must be caused by values that one has acquired in the right, non-freedom-undermining way. If one’s behaviour is caused by values one acquired in a different manner, one is unfree and non-responsible.⁷⁸

A major problem with the above kind of compatibilist theory is that it is usually argued for by appealing to intuitions about various imagined cases. A scenario is described, and then introspection of one’s intuitions is supposed to provide an answer to whether the main character is morally responsible for what he did. However, using the content of one’s intuitions regarding thought experiments as an argument for a thesis is problematic for many reasons. It is not at all obvious that our intuitions show us the truth of the matter. Even if it were the case that some philosophers have truth-tracking intuitions, it cannot be true about everybody working in the field, since intuitions differ significantly. Naturally, everybody will think of his own intuitions as reliable and differing intuitions as non-reliable, but there is no rea-

⁷³ As I pointed out earlier, most of the free will discussion focuses on moral responsibility rather than my wider notion of moral agency.

⁷⁴ Fischer and Ravizza, 1998, p 38

⁷⁵ *ibid*, p 48

⁷⁶ *ibid*, p 40, “the normal mechanism of practical reasoning” sounds like it might be mental.

⁷⁷ *Ibid*, p 170 and onwards

⁷⁸ Mele, 2001, pp 144-176

son to suppose that one person's intuitions are better than another's.⁷⁹ Historically, it has been common to accuse one's opponents of having had their intuitions twisted by too much philosophy and argue that the man in the street has natural, unsullied intuitions that correlate with one's own. Once some philosophers started to investigate man-in-the-street intuitions rather than just speculate about them, this line of argument became more difficult to sustain; it turns out that laypeople differ as well.⁸⁰

Thus, one advantage that PPC has over causal compatibilist theories is that it does not rely on intuitions about thought experiments, but has PDR as its main argument.⁸¹ Another advantage is that PPC is unthreatened by epiphenomenalism.

It is often assumed that many human beings uncontroversially fulfil various compatibilist conditions for having the morally relevant kind of freedom, and the controversial question is whether we also fulfil the libertarian conditions or not. That is not so. Any compatibilist who thinks it a necessary freedom condition that intentions, desires or the like cause a person to act, presuppose that it is possible for mental phenomena to cause physical events. This *is* controversial. Jaegwon Kim argues that physicalism, the thesis that all mental phenomena can be reduced to physical ones, has to be true for mental causation to be possible.⁸² Without physicalism, there would be no

⁷⁹ I know this is a controversial claim. Some philosophers do think that one can distinguish between, on the one hand, intuitions that just reflect some instincts that evolution shaped us to have, prejudices, and irrational emotions, and on the other hand, intuitions that track the truth. I do not think that is possible; the most one can do is to distinguish those intuitions that survive a process of critical reflection from those that do not, and widespread intuitions from not so widespread ones. I suppose that all philosophers engaged in the free will debate have done much reflecting over their moral responsibility intuitions already. They still have their differences, and there is no overwhelming majority for one view over another.

⁸⁰ See, for instance, Nichols and Knobe, 2007; Nelkin, 2007; and Vargas, 2006.

⁸¹ Obviously PDR, as any philosophical argument, ultimately relies on intuitions as well (although not intuitions built on thought experiments). There is probably no getting around that. However, I think any intuition that lies behind PDR is less spurious than the intuitions we have regarding the moral responsibility of some agent in a far-fetched, complicated scenario that does not resemble anything we have ever encountered in real life.

⁸² Kim, 2006, p 300

agency, according to Kim – and this would imply no moral agency either. On the subject of substance dualism he writes:

As will be recalled, the fourth component of Descartes' dualism is the thesis that minds and bodies causally influence each other. In voluntary action, the mind's volition causes our limbs to move; in perception, physical stimuli impinging on sensory receptors cause perceptual experiences in the mind. This view is not only commonsensical but also absolutely essential to our conception of ourselves as agents and cognizers: Unless our minds, in virtue of having certain desires, beliefs, and intentions, are able to cause our bodies to move in appropriate ways, how could human agency be possible? How could we be agents who act and take responsibility for our actions?⁸³

Kim thinks that physicalism has to be of a reductive type-type identity kind. Donald Davidson has argued that anomalous monism, a theory according to which there is a token-token identity between mental and physical events, is enough to secure mental causation,⁸⁴ while Kim thinks this theory eventually faces epiphenomenalism problems as well.⁸⁵ I think Davidson is right, but that claim is highly controversial. In any case, in philosophy of mind the question of whether the mental is epiphenomenal or not is a real controversy. A compatibilist theory according to which free action is action caused by beliefs, desires, the appearing of a reason or the like, implies that we can only act freely given that epiphenomenalism is false. It is thus false to say that only libertarianism requires the truth of controversial metaphysical theories – some compatibilist theories do as well.⁸⁶

⁸³ *ibid*, p 40

⁸⁴ Foster and Swanson, ed, 1970, pp 79-101

⁸⁵ Kim, 2006, pp 188-189

⁸⁶ In Fischer, 1994, p 174, Fischer argues that a morally responsible choice is one where the action issues from a biological mechanism that is responsive to reasons. Here the mental/physical problem is how *reasons* can influence a physical mechanism – at least it is a problem if we assume that reasons are not identical with something physical. A physical mechanism can be influenced by other physical things, but if a reason is something metaphysical or mental we have a problem that needs to be solved.

Mele has escaped this problem by stating that it is enough that the “physical correlates” of intentions cause one’s actions.⁸⁷ This means that the question of mental causation is no problem to Mele. It would still be a problem for him if neurologist Benjamin Libet was right in his claim that actions are caused by neural events that are *not* simultaneous with intentions, or if psychologist David Wegner was right in that neither intentions nor their physical correlates play a causal role in producing action, and Mele has therefore argued at great length against Libet’s and Wegner’s interpretations of their experimental data.⁸⁸ I do think Mele’s arguments against them are successful. Still, on PPC none of this matters at all. PPC is simply immune against any arguments about how our brain works, since it implies that the causes of one’s actions are not important in themselves. Practical freedom is not just compatible with the truth of either determinism or indeterminism, but also compatible with the truth of epiphenomenalism, as well as with the truth of Libet-type theories. From a practical perspective it makes no difference whether my decisions cause my body to move, whether decisions are merely some kind of epiphenomenological dust produced by my brain as a side-effect of neurological events, or whether the neurological causes of my bodily movements are some brain events other than the physical correlates of my mental intentions and decisions. It would still be true that I have to choose what to do in various circumstances.

Therefore, PPC has the virtue of making moral agency compatible with determinism, indeterminism and even epiphenomenalism, and it does not rely on intuitions elicited by far-fetched thought experiments. However, it does build on the controversial Kantian idea of there being “different perspectives” from which to view the world – one theoretical, and one practical. Some readers have probably been thinking, on reading section after section

⁸⁷ Mele, 2009, p 11

⁸⁸ *ibid*, pp 49-90

on how we “choose” between different “options” that we have in a “practical perspective”, that this just means that even determined agents *imagine* themselves to choose and have options. Some readers may think, in short, that what we see when we view the world from what I have called a “detached, theoretical perspective” is really the *truth*, according to which there may be no options to choose between.

In the following chapter I will show that there is a non-mysterious way to understand this talk of perspectives that does not presupposes Kantian idealism, nor makes the practical perspective a mere illusion.

4. Different Perspectives

4.1. Different Perspectives – Contradictory Beliefs?

Kant famously argued that we have free will in a much deeper sense than just being able to do what we want, despite the world being deterministic. Given Kant's view on determinism, this is not particularly surprising. In brief, Kant thought that the basic characteristics of the empirical world, the world of experience, depend on our mental apparatus. Since we can only experience things as taking place in time and in a three-dimensional space, the empirical world has one dimension of time and three of space. How things are in themselves, independent of our experience of them, we can never know; but we can know with absolute certainty that we will never experience the discovery of some region of the world where, for instance, there is no time. Now cause and effect are, according to Kant, prerequisites for experience just like time and space. Kant equates cause and effect with determinism; he consistently writes about causes as necessitating their effects, rather than, for instance, making them probable.⁸⁹ The empirical world is thus deterministic because this is the way we experience things; but once again, how the world is in itself, regardless of our experience of it, is a different matter.⁹⁰

John Perry distinguishes between “strong” and “weak” accounts of laws of nature. According to Perry, compatibilism is a tenable position when

⁸⁹ See for instance Kant, 2004, pp 185 and 199.

⁹⁰ Kant, 2004, pp 275-276

combined with a weak account of laws.⁹¹ Laws of nature in the strong sense are something *more* than a mere tool for prediction, and more than just regularities and patterns in events; they are what *create* the regularities and patterns, a substantial force (for want of a better word) in the universe. Since Kant believed that determinism lies in the way we experience things rather than in the world as it is in itself, his “laws of nature” seem to be of the weak kind.

I will, however, follow Kantians like Korsgaard and Bok and not base my arguments on any particular understanding of what laws of nature are, or how to understand determinism. Let us suppose, for the sake of argument, that laws of nature are strong and the world is mechanistic,⁹² regardless of whether there is complete determinism or some randomness in the system. With this assumption in place, one might worry that “practical freedom” just is a fancy name for the mistaken, but perhaps unavoidable, *belief* that we can choose what to do out of several different options, when in reality our actions are just products of the vast mechanism we call the universe. If this were the case, PPC would seem to be terribly counterintuitive. Why would having a *mistaken belief* ground moral agency and responsibility? The worry that the freedom we experience from a subjective or practical perspective is really just an illusion has been examined at some length by Nagel in his *The View from Nowhere*, as well as Dana Nelkin in her article “Two Standpoints and the Belief in Freedom”.

Nagel writes that when we try to learn more about the world, we usually do this by trying to get a more objective picture of it. To explain what he means by “objective” Nagel writes:

⁹¹ Campbell, ed, 2004, p 231

⁹² As Bok uses the word, meaning that human actions either are caused deterministically the same way as every other event, or is subject to the same kind of randomness as every other event.

To acquire a more objective understanding of some aspect of life or the world, we step back from our initial view of it and form a new conception which has that view and its relation to the world as its object. ... The old view then comes to be regarded as an appearance, more subjective than the new view, and correctable or confirmable by reference to it.⁹³

A view or form of thought is more objective than another if it relies less on the specifics of the individual's makeup and position in the world, or on the character of the particular type of creature he is.⁹⁴

For example, I may see a red object. When I take a step back and acquire a more objective perspective, I realise that my visual experience was caused by light-waves of a certain wavelength that this object reflects, and which then affect my eyes and brain in a certain manner. The idea that this object is red is subjective in the sense that it depends on how my senses work. An alien creature who felt light-waves by special receptors on his skin would perhaps describe the surface of the object in terms of pressure instead of colour. When it comes to colour we may disagree, but as long as we are rational and have some way to examine physics we can come to agree that this object does indeed reflect light-waves of a certain wavelength. To say that an object reflects light-waves of a certain wavelength comes closer to the objective truth of the matter than just saying that it is red.⁹⁵

In general Nagel thinks that a progressively more objective perspective is the best and most fruitful way to enlarge our understanding. There are some facts that cannot be found when looking at the world from an objective standpoint though – facts about *what it is like* to be somebody or experience something. I may, for example, look at the brain of somebody who sees something red, and describe everything that happens in there, but I would

⁹³ Nagel, 1986, p 4

⁹⁴ *ibid*, p 5

⁹⁵ *ibid*, p 14

not have captured *what it is like* to see red in my description. Yet there are facts about what it is like to see red; experiences are part of reality.⁹⁶ A more objective standpoint thus often reveals a truer picture of the world, but some parts of reality drop out of sight, which means that an objective standpoint is not always the right one to take up when one wants to learn about the world. If I think and feel, then there is a certain phenomenology from the inside, the *experience* of what it is like to think and feel. From a more detached and objective perspective there are only various events in the brain. If I choose my actions, there is a certain phenomenology from the inside, the experience of freedom. From an objective perspective there are once again only events in the brain. However, the analogy between mind and freedom is mine; Nagel himself does not consider freedom and mind analogous in this respect. He thinks that mind is real, although it drops out of sight when we take an objective perspective on things, while freedom is revealed as an illusion by that same perspective.

Nagel writes that we experience that we have many open alternatives before us, and can choose between them, and we experience our choices as neither determined nor random. When we take up a more objective perspective, however, we see that this is impossible. We cannot *possibly* be free in *that* sense. We want to know as much as possible about ourselves, so that we can make more informed and therefore more autonomous choices, but when we know enough about ourselves we realize that autonomy is impossible.

It is of course common among philosophers to discuss how we feel, think or experience things, and it is very common in the free will debate. The problem is that there is no universal agreement on what our experiences are like (possibly because people are different). As I wrote in Chapter 2, Kapitan thinks that the feeling of freedom agents have is simply a feeling of having an efficacious will and an epistemically open future, while Nagel thinks that agents feel an actually impossible power within themselves. For my own

⁹⁶ *ibid* p 15

part, I cannot recall having any particular experience of being a *causa sui* in the impossible sense that Nagel seems to experience on a regular basis, although I constantly experience that I choose things and that they are up to me in some sense. Alfred Mele described this well, so I will quote him:

Here I am, sitting at my computer, composing this paragraph. I am also a bit thirsty. I now have the experience of entertaining two genuine options, two alternative courses of action open to me: one is to continue to sit here thinking and typing, for several minutes at least; the other is to walk down the hall to the drinking fountain in a few minutes. I consciously regard these options as things I can do and as things I may or may not do, as I please. At least in that sense, I have an experience of it being “up to me” which of these I do, or an experience describable in that way. However – and this is just a report on how things seem to me just now – I find in myself no experience that the total state of myself and the world now is causally compatible with my sitting here typing and thinking for several minutes and causally compatible, as well, with my instead going to the drinking fountain in a few minutes. Nor do I find in myself an experience of regarding things this way. Call these last two experiences – experiences that it seems to me I lack – *deep openness* experiences.

How can I have my “up-to-me” experience without having a deep openness experience? Well, perhaps I am not a very deep fellow along experiential lines.⁹⁷

Perhaps I am as shallow as Mele experience-wise, but just as he does, I often experience that I can do different things as I please, that I have practical freedom – this *deep openness* is still not something I recognise. Nagel, however, experience that he has an impossible kind of libertarian free will.⁹⁸ The

⁹⁷ Mele, 2001, pp 135-136

⁹⁸ I think it would be wrong to say that libertarian free will is impossible per se. Event libertarian theories like Kane’s or Mele’s “modest libertarianism” (Mele, 2001, p 211) are called “libertarian”, and although it can be debated whether they give us all the freedom that could be worth wanting, or if they give us more than compatibilist freedom can offer, they are not incoherent or impossible. Pereboom does not think agent causation libertarian free will is incoherent either, only that the existence of agent causation is not probable given our best

impossibility of this kind of freedom is just not evident unless he takes up a more detached perspective. However, Nagel's "freedom from a subjective perspective" or "deep openness" is not the same thing as practical freedom, as one can have the latter without the former.

To illustrate the difference an analogy could be useful. Suppose that I merely experience a pain in my stomach. If it turns out that the pain was caused by intense stress, the pain is still real. Suppose on the other hand that I experience the presence of an alien object in my stomach that gives me pain. If it turns out that there is no alien object there, only stress, the alien object was an illusion. Practical freedom is like the first example, and the "freedom from a subjective perspective" that Nagel discusses is like the second. It may very well be the case that the past and the laws of nature, or some quantum events in my synapses, make it the case that I believe that I am self-aware and believe I have an efficacious will, and this in turn means that I must choose what to do. These facts about me may not look like *freedom* from a detached theoretical perspective, but they do not *disappear* when I view things this way. It is not like an objective standpoint will reveal that there is actually no connection between my decisions and my bodily movements, that I cannot think in terms of reasons, or that there exists an alternative to choosing my actions.⁹⁹ Nagel's "freedom from a subjective perspective" is like the second example. If Nagel feels that he has the power to cause things to happen in a way that is neither caused nor random, a more objective standpoint might show that Nagel cannot have the power he thought he had.

In short: "Practical freedom" is a term that stands for capacities that the agent has regardless of which perspective one takes, although they may not

scientific theories (Pereboom, 2001, pp 415-416). That is a controversial claim of course, but perhaps he is right. Nagel, however, thinks he experiences something which *is* impossible.

⁹⁹ "Choice" understood as a mental event, as thinking something like "Okay, I'll do *this* and not *that*." If "choice" is interpreted the way it is used in the consequence argument (van Inwagen, 1983, s 69) it may of course be proven impossible by theoretical arguments, but that is not how I use it when discussing practical freedom.

look like *freedom* from a theoretical perspective. Nagel's "freedom from a subjective perspective" stands for imagining that one has certain capacities that one really lacks, which is a very different thing.

Dana Nelkin has more recently raised a similar concern.¹⁰⁰ It is common among neo-Kantians to say things like "I believe that I am free from a practical standpoint, but not from a theoretical standpoint". Nelkin argues that if I believe both the proposition "I am free" and the contrary proposition "I am not free", then I believe a contradiction, and I am irrational. If it is somehow necessary for me to believe the former proposition when engaging in practical reasoning and the latter when engaging in theoretical, it just goes to show that it is necessary for me to be irrational, but I am irrational all the same. To claim simply that I believe these contradictory statements from different standpoints does not resolve the issue, since it is unclear even what it means to believe something "from a standpoint".¹⁰¹ The only plausible interpretation of this phrase that Nelkin can come up with points to some kind of voluntary irrationality,¹⁰² or cases where it may in some sense be "rational to be irrational", but the irrationality is still there.¹⁰³

I think the solution lies in a contextualist interpretation of these claims. Perhaps "freedom" in a practical context means the ability and necessity to choose what to do, while "freedom" in a theoretical, scientific or causal-explanatory context either means nothing at all, or means some almost god-like power.¹⁰⁴ To say that I am free from a practical perspective but not free

¹⁰⁰ Nelkin, 2009

¹⁰¹ *ibid* p 569

¹⁰² *ibid* p 573

¹⁰³ *ibid* p 574

¹⁰⁴ This is a controversial claim, of course. Daniel Dennett, 1981, pp 233-255 for example, has argued that it is impossible to adequately explain people's behaviour without using mental terms like "intention" and the like. One cannot reduce everything to physics. One might be able to explain all bodily movements by physics, but certain patterns would fall out of sight if one did. He also thinks that this is relevant for compatibilism – people's behaviour is not solely events caused by the past and the laws of nature, but also free actions performed for reasons. If Dennett is right, then Kant (as well as Korsgaard and many other Kantians) was

from a theoretical perspective, would then mean that I am free if the word is interpreted in a sense that is relevant to practical thinking, but not free if the word is interpreted in a sense relevant to theoretical thinking, to science and causal explanations. There is of course nothing irrational about this.

4.2. A contextualist Interpretation

Contextualism is largely the thesis that the meaning of certain words (like “tall”, “expensive” or “local”) depends on the context in which said word is uttered. For instance, Korsgaard writes that freedom is something found in the first-person perspective of deliberation, not in the explanatory perspective scientists take.¹⁰⁵ A possible contextualist interpretation of this claim would be that in a certain context, that of deliberation, “freedom” refers to the ability to choose one’s actions, while in another context, that of science, “freedom” might refer to an absence of causal necessity, or perhaps to nothing at all. This subsection will provide a contextualist way to understand the idea of a practical versus a theoretical perspective. It is not meant as an exercise in exegesis of the previously discussed neo-Kantians, but as a plausible interpretation, one I will rely on when I write about practical and theoretical perspectives in my own theory.

4.2.1. Lewis-style Contextualism

I am not the first philosopher to come up with an explicitly contextualist theory of freedom. John Hawthorne has suggested that contextualism might be a good approach to freedom in his article “Freedom in context”, in which

wrong in thinking that “freedom” cannot refer to anything we can find in an explanatory, scientific context. I do not find these arguments convincing. I agree that some important patterns would probably fall out of sight if we described the world in completely physical terms, but even if humans are described in psychological terms they may seem governed by laws rather than free. *If* Dennett is right though, compatibilism is true, just like I claim – although my entire discussion of perspectives and contexts would be superfluous.

¹⁰⁵ Korsgaard, 1996a, pp 94-96

he analyses freedom in a way analogous to David Lewis' contextualist analysis of knowledge,¹⁰⁶ and Steven Rieber has made further developments on this kind of theory.¹⁰⁷ However, I think Lewis' theory has some problems that transfer to Hawthorne's and Rieber's, which a good contextualist theory on freedom should strive to avoid.

I think the problems with Lewis-style contextualism stem from the fact that Lewis' theory is primarily a semantic one, intended to solve the following problem: We know many things. I know that I am a married woman, that I live in a wooden house, that I own a blue car and have four dogs. I know that the Earth revolves around the sun, and that each year the movements of tectonic plates bring Europe and North America slightly further apart. However, I would not say that I *knew* these things if, for instance, fake-wood was a common material for houses and I had not bothered to check that my house was made of real wood before I bought it, or if I was colour-blind in the blue spectra and could not distinguish a blue car from a purple one. I only say that I *know* I live in a wooden house and own a blue car because other possible explanations of my experiences have been ruled out. It seems that knowledge must be infallible, that one cannot know anything unless all other possibilities are ruled out. However, regarding my supposed knowledge of the facts listed above, there are alternative possibilities that I have not ruled out, which are consistent with *all* the evidence I can have for those claims. Perhaps, for instance, I am not a married woman with four dogs, but a brain in a vat. It is absurd, Lewis claims, to deny that people have a great amount of knowledge of ordinary, empirical matters, as those I listed above, and yet there is a very good argument to the effect that such knowledge is impossible; knowledge requires ruling out all other possibilities, and it cannot be ruled out that I am a brain in a vat, deceived by a demon or the like.¹⁰⁸

¹⁰⁶ Hawthorne, 2001

¹⁰⁷ Rieber, 2006

¹⁰⁸ Lewis, 1996, p 549

Lewis' solution to this puzzle is to claim that "I know P" means that I have ruled out all other possibilities – except those possibilities that I *properly ignore*. If somebody asks me how to reach the highway from the country road and I tell him, I properly ignore the possibility that my memory of the roads of this area might have been created afresh five minutes ago by a deceitful demon, and might not correspond to the real roads at all. If I deliberate about which charity to support, I properly ignore the possibility that I might be the only person in the universe and everybody else just part of my hallucination. Lewis lays down some rough rules for what can be properly ignored, which I need not go into here, but the above cases are obvious instances. If this is what "I know P" means, then people often tell the truth when they claim to have knowledge of empirical facts. Still, as soon as a sceptical argument is brought up and heard it is no longer ignored, neither properly nor otherwise.¹⁰⁹ This is why Lewis called his paper "Elusive knowledge" – even if one had knowledge to start with, that knowledge would dissipate as soon as somebody mentioned a deceitful demon or any other equally far-fetched sceptical alternative.¹¹⁰

Hawthorne and Rieber suggest that something similar holds true for freedom. An action is free if it was not caused by any factor outside the agent's control – with the exception of such causes as are properly ignored. In most contexts we properly ignore the distant past and the laws of nature, and instead look to more immediate causes such as, for example, a drug put in the agent's drink (outside his control and thus freedom-undermining) or the agent's normal beliefs and desires (under his control, and thus not freedom-undermining). However, as soon as "the distant past and the laws of nature" are brought into discussion, freedom evaporates.^{111 112}

¹⁰⁹ *ibid*, p 559

¹¹⁰ *ibid*, pp 560 and 566

¹¹¹ Hawthorne, 2001

¹¹² Rieber, 2006, 228

Hawthorne's and Rieber's contextualism has not really swayed the free will debate. Compatibilists and incompatibilists of the traditional, non-contextualist kind still dominate the discussion, probably because Hawthorne's and Rieber's theories are semantic in nature, while semantics is not what primarily drive most of the free will debate. Just like Lewis' theory, theirs allow sceptics to make common-sense judgements on the truth-value of ordinary statements regarding free actions or knowledge, without giving up on the basic scepticism, but some important worries remain.

Hawthorne briefly considers that moral responsibility, or "accountability" as he puts it, might be analysed in a contextualist manner as well.¹¹³ Such analysis would have absurd implications. According to Hawthorne's theory, philosophers, being aware of the sceptical arguments, will lie if they say somebody acted freely on any occasion, while laypeople will often tell the truth. If one is morally responsible for actions that are "free" in this contextualist sense, this leads to the unhappy conclusion that laypeople are often right when they hold others responsible for what they do, but philosophers never are. This is absurd. If George is morally responsible for driving into and destroying both my and my neighbour's mailbox with his car, this has certain implications; for instance, both my neighbour and me would be justified in blaming George, and it would be *fair* that he payed for having our mailboxes replaced. Hawthorne suggests that it might be the case that my neighbour is fully justified in blaming George and having him pay for her mailbox, but I am not, simply because I happen to be a philosopher. Rieber wants to leave the question of moral responsibility open; the concept of moral responsibility may be either compatibilist, incompatibilist or contextualist.¹¹⁴ However, Rieber does note that it "certainly sounds odd" to say that "John is morally responsible for his action, but his action was not free".¹¹⁵ Yet this must sometimes be true if "free" is context-dependent while "moral

¹¹³ Hawthorne, 2001, p 70

¹¹⁴ Rieber, 2006, pp 241-242

responsibility” is non-contextualist compatibilist. If “moral responsibility” on the other hand is non-contextualist incompatibilist, the sentence “John acted of his own free will and did exactly what he wanted to do [further conditions could be added if need be], but he was not morally responsible for his action” will sometimes be true, which seems quite absurd as well, at least unless we have already proven that utilitarianism or some other ethical theory that simply does not make use of the concept “moral responsibility” is true. Finally, the problems that follow from assuming that moral responsibility is contextualist in the same way as Hawthorne and Rieber suggest that freedom is, have already been noted. This is one problem for this kind of Lewis-inspired contextualist free will theory; absurd consequences for moral responsibility will follow whether that concept is also context-dependent or not.

The second problem for this kind of contextualism is that if one considers knowledge important, it would be natural to desire a knowledge that is not elusive but stable. Although Lewis argues that our everyday knowledge deserves the name, it seems like some kind of pseudo-knowledge since it vanishes as soon as somebody says “brain in a vat”. Similarly, if one considers freedom important because it is necessary for moral agency and moral responsibility, one might desire a stable kind of freedom that survives scrutiny and awareness of all the facts.

Richard Feldman, in “Freedom and Contextualism”, criticises Hawthorne on a number of points. For instance, philosophers often see random chance as a threat to freedom at least as severe as determinism, but Hawthorne only discusses causation as a threat to freedom.¹¹⁵ However, this problem with Hawthorne’s article is easily remedied. Just add that an action, to count as free, cannot be random, except for such randomness as we properly ignore. If there are some random quantum events in the agent’s brain for example,

¹¹⁵ Rieber, 2006, p 241

¹¹⁶ Campbell, ed, 2004, p 272

they are usually properly ignored. The main problem according to Feldman, and I agree, is that this kind of contextualist theory hardly deserves the name “compatibilist” which Hawthorne wants to give it. Even if the theory allows that the statement “he acted of his own free will” can be truly spoken by somebody living in a deterministic universe, it claims that as soon as, for example, the consequence argument is mentioned (and the past and the laws of nature thereby invoked), then it is true to say that nobody ever does anything freely.¹¹⁷ Normally we think of compatibilists as disagreeing with incompatibilists when the latter claim that determinism makes freedom impossible. A Hawthorne “compatibilist” cannot do that, as Hawthorne himself notes,¹¹⁸ since he accepts the incompatibilist arguments as soon as they are raised.

Fortunately I see no reason to accept the particular version of contextualism offered by Hawthorne and Rieber. It is possible to come up with a much better contextualism that avoids these serious problems.

Lewis makes a point of it being impossible to ignore, for example, the possibility of being a brain in a vat, if somebody mentions it. Lewis writes:

When we say that a possibility is properly ignored, we mean exactly that; we do not mean that it *could have been* properly ignored. Accordingly, a possibility not ignored at all is *ipso facto* not properly ignored. What is and what is not being ignored is a feature of the particular conversational context. No matter how far-fetched a certain possibility may be, no matter how properly we might have ignored it in some other context, if in this context we are not in fact ignoring it but attending to it, then for us now it is a relevant alternative.¹¹⁹

¹¹⁷ Campbell, ed, 2004, p 271

¹¹⁸ Hawthorne, p 73

¹¹⁹ Lewis, 1996, p 559

Is resistance useless? If you bring some hitherto ignored possibility to our attention, then straight-way we are not ignoring it at all, so *a fortiori* we are not properly ignoring it. How can this alteration of our conversational state be undone? If you are persistent, perhaps it cannot be undone – at least not so long as you are around. Even if we go off and play backgammon, and afterward start our conversation afresh, you might turn up and call our attention to it all over again.¹²⁰

Likewise Hawthorne suggests that one cannot ignore the past and the laws of nature when they are mentioned,¹²¹ and the same assumption lies implicit throughout Rieber's paper. This is, of course, true *in a sense*. If somebody says "suppose you're a brain in a vat" I cannot help but respond to these words in some way, perhaps by picturing a brain in a vat, or at least having a brief thought of brains in vats. If "ignoring" a spoken sentence means that it has no effect on my mental life whatsoever, then I cannot ignore it. In the same (rather odd) sense of "ignore" it is impossible to ignore a person if he stands within my field of vision. However, even if I notice that a person stands at a certain spot in my field of vision it is of course possible for me to ignore him in another sense – I may consider him uninteresting, irrelevant or not worth paying any real attention to. In this latter sense of "ignore" one can perfectly well ignore the brain-in-a-vat possibility and all kinds of argument for scepticism even though they are brought up. If a colleague of mine protested against this text that I might be wrong about Lewis' theory, I would pay attention, since it is important to be accurate when referring to other philosophers, and it is just possible that I have misunderstood something. However, if my colleague went on to say that a hallucination-inducing demon might deceive me every time I read an article or a book by David Lewis, I would happily ignore him, and I daresay most philosophers would do the same. Hawthorne consistently uses the word "ignore" in Lewis' slightly

¹²⁰ *ibid* p 560

odd sense, rather than the more everyday sense in which it means something like “not caring about”. However, in a footnote Hawthorne says that an alternative to his and Lewis contextualist theories would be one in which the propriety of ignoring is what varies from context to context, and that it might be proper in some contexts to ignore certain possibilities/causes whether they are actually brought to attention or not.¹²² This is what Bok seems to have in mind, when she writes that the concept of “possible action” serves different functions in theoretical and practical reasoning. The difference between them is not that agents, as a matter of fact, do not think about the past and the laws of nature when engaged in the latter activity; it is that agents have a good reason not to focus on distant causes of their actions when deliberating, since focusing on the causes cannot help them decide what to do.¹²³

4.2.2. Williams-style Contextualism

Epistemologist Michael Williams has a contextualist theory that treats the propriety of ignoring as variable – Williams consider certain possibilities to be irrelevant, to be such that we *should* ignore them, in certain contexts, whether actually brought up in discussion or not. A contextualist analysis of the freedom theories I have discussed in this dissertation would rather be analogous to Williams’ than to Lewis’ theory of knowledge. Williams’ motivation for contextualism is that it is strange to suppose that there is such a thing as *the* standard for what counts as knowledge that differs from the standards used in everyday life or by various empirical sciences but still binds us all somehow. Where would such a standard come from, what would it be, and how could it bind us? If this idea of a standard for all knowledge is rejected, some kind of contextualism becomes an obvious alternative. Wil-

¹²¹ Hawthorne, 2001, pp 68-69

¹²² Hawthorne, 2001, p 79

¹²³ Bok, 1998, p 105

Williams then considers the traditional contextualism that Lewis represents as problematic, since it takes into account only high and low standards for what counts as knowledge (and according to this theory the highest standards are set by epistemology where the sceptical arguments come in). Bringing up new possibilities of error in the discussion is, on Lewis theory, always equivalent to raising the bar. Williams on the other hand writes that there is an important difference between *raising* the standards for knowledge on the one hand, and *changing perspective* and viewing things from another angle on the other hand.¹²⁴ He argues that the epistemological standard for knowledge *differs* from many other standards, but it is not a *higher* or *superior* standard, in the way that standard for historical knowledge among scholars is higher than the standard for schoolboys or quiz contestants.

If a school boy in a classroom, or a contestant in a quiz show, is asked what date Columbus arrived in America, it will be generally considered that he *knew* the answer if he has read the date in a history book and remembers doing so. Suppose, on the other hand, that a respected historian has published a paper in a prestigious scholarly periodical arguing that Columbus actually arrived on October 11, 1492, as opposed to the generally accepted date of October 12 that same year, and presents good evidence for his claim. If I am a historian myself and want to argue against the first historian, it is obviously no good to say that I know it was October 12 because I read it in a history book. In this context, the standards for knowledge are much higher than among school children or quiz contestants.¹²⁵

Williams and other epistemologists writing on this subject seem to take it for granted that the standards in this situation are higher, not just different, but why that is may require an explanation. Here is mine: The scholar as well as the quiz contestant and schoolboy ultimately rely on archaeological evidence and old documents from the historical period. The scholar does not

¹²⁴ Williams, 2001, p 160

¹²⁵ Ryseview, 2011, section 1.

rely on any special presuppositions that the schoolboy and quiz contestant does not share; the difference between them is simply that the scholar has to examine archaeological evidence and old documents *himself* and produce *new* knowledge from these sources.

Williams thus agree that standards can be raised from one context to the other, but he does not think that bringing up sceptical epistemological concerns is to further raise the standard. If I am engaged in historical research, I should of course be scrupulous and careful, rather than trust any book I happen to come across. However, if I seriously doubted whether the Earth even existed five minutes ago, if I seriously considered whether it might not be the case that the whole world was recently created complete with fake historical documents and fake memories in all individuals, this would not result in a particularly scrupulous approach to history. This would not mean, as Lewis implies, that I had raised the standard for historical knowledge even further. On the contrary, it would mean that I was no longer engaged in historical research at all. According to Williams, it is proper to ignore these possibilities when studying history, since ignoring them is a necessary part of the project of history.¹²⁶ If I go back to my own explanation of why the standards of knowledge count as raised rather than just changed when we go from school children to scholars, I can explain why the standards between history and epistemology would count as changed by the same line of reasoning. Epistemologists do not rely on the same kind of evidence as historians; archaeological evidence and old documents are irrelevant to them. Epistemologists disregard all presuppositions historians have, while introducing completely different presuppositions of their own (like, for instance, what Anders Wedberg has called “the immanence idea”, when it comes to brain-in-vat worries – the idea that the human mind is somehow locked up inside itself, and cannot know anything directly about the outside world¹²⁷). It is not

¹²⁶ Williams, 2001, p 160

¹²⁷ Wedberg, 1966, p 15

that epistemologists have a *superior* standard of knowledge compared to historians, they simply have a *different* one, relevant for different purposes.

The reader should now have a good grip on William's contextualist epistemology, and how it differs from Lewis' theory. Let us now return to the question of free will, and how the theory of practical freedom can be interpreted in a contextualist sense along the lines of Williams' theory.

We have seen that Williams' epistemology differentiates between raising the standard for knowledge, and switching to a different context where the standards are simply different. Practical perspective compatibilism does the same. We could discuss freedom from a causal perspective and define a free action as one that is caused in the right way. We can then set a rather low standard for freedom and consider all actions that are immediately caused by my desires rather than something external to me free, or we can set a high (perhaps impossible) standard and require that my actions must not have been caused by anything outside me no matter how far back we go in the causal chain. However, if we switch to the context of deliberation and choosing what to do, formerly relevant factors about the past and the laws of nature become irrelevant. Not because we have lowered the standards, but because we have *changed* standards when we changed perspective from causal explanation to deliberation. Many factors that play a crucial role in causal explanation are simply irrelevant for deliberators. That a blue car drove down the main street of Örebro on 20 November 1969 may be part of a complete causal explanation of why I chose to become a philosopher, but it could only be a reason for choosing this profession given some very complicated back story. That I love philosophy may be part of a full causal explanation as well as figure as a reason for choice. When taken as a reason, however, it is irrelevant whether there are some complicated psychological law that determine everyone with my exact pattern of preferences and personality traits to become a philosopher, although this would be highly relevant for

somebody looking for a causal explanation. It is, however, relevant whether I think it is rational to choose a profession based on what one loves (I could have thought, instead, that one should choose a profession solely based on considerations of the job market and one's talents). Although some factors may figure in either a causal explanation or as reasons in deliberation, causes *considered as causes* are irrelevant for deliberation. Trying to find out whether my love of philosophy caused me to do this or that could be an interesting exercise in introspection, but no decision will come out of it. To choose a philosophy career I must consider whether a love of philosophy constitutes a good reason to choose this job or not.

Korsgaard came up with a thought experiment to illustrate vividly how causes considered as causes are irrelevant for deliberation.¹²⁸ Suppose that I have volunteered as test subject for an experiment, where scientists will remote control all my thoughts and actions for one day. This does not entail that I will be reduced to a puppet, with arms and legs twitching involuntarily in response to the scientists' pushing of buttons. Neither does it mean that I will try to do one thing, but feel an irresistible urge to do quite another thing that the scientists decided for me. Instead, the scientists will work through my usual mental mechanisms in order to produce deliberations and intentions that issue in actions. Now, let us see how this experiment affects my practical freedom: The day of the experiment arrives. I wake up, have breakfast, pick up my computer, and do some work, just like any other day. Suddenly the thought strikes me that it is *silly* to work, as I only do it because the team of scientists cause me to. I rebel; I decide to skip work to go shopping. On my way to the mall though, I realise that the scientists caused me to firstly feel silly, secondly rebellious and thirdly to go shopping. Everything feels pointless. I decide to just sit down and stare. But then I realise that sitting and staring is no better than shopping. Sitting down to gaze listlessly in front of me will have been just as much caused by the scientists as shopping

or working, the only difference being that as long as I sit down and stare I neither get any work done nor new clothes to wear. That the scientists are controlling me is a fact that could give me reason for action given some particular assumptions; suppose, for instance, that I had promised to call my supervisor if scientists gained control of me. Given this background, being controlled by scientists gives me reason to call him. We might also claim that Korsgaard's example is a bit underdescribed, since she does not explicitly tell us that the scientist will respect my "character" (although the fact that they will produce decisions through the usual mechanisms could be interpreted this way). If my philosophical reasoning was different from normal during the day of the experiment, this may count as a reason not to work, but rather do something more inconsequential. Suppose, however, that the scientists work under certain constraints, and will cause me neither to philosophise in a way I would not normally nor to buy clothes I would normally abhor. Then, barring special cases like the one where I had promised to call my supervisors, the fact that I am controlled by scientists cannot tip the scales in favour of one action over another; it will simply be irrelevant when deciding what to do. Even with the scientists in their control room playing around with my beliefs and desires, I have my practical freedom, since I must still make up my mind and choose my actions.¹²⁹

This also answers Nelkin's question of why one should not make use of the theoretical belief in determinism in practical deliberation, since there is no general reason not to use theoretical beliefs as data when trying to make up one's mind.¹³⁰ In the above example, where I deliberated about whether to work or shop, theoretical facts about the state of my bank account, my wardrobe, my salary and my career could very well be relevant, because these are the kind of facts that can count in favour of one considered option rather than the other. However, determinism or being controlled by scientists is

¹²⁸ Korsgaard, 1996b, pp 162-163

¹²⁹ Korsgaard, 1996b, pp 162-163

irrelevant, because the fact that my values and thoughts were determined or created by scientists, as opposed to having a different history, cannot influence my reasons for action, as seen above. They cannot help me make up my mind by tipping the scales in favour of either working or shopping.

I do not argue that we are free to choose what to do because our choices are not caused by anything outside our control. I argue that we are free to choose what to do because those causes of action that lie outside our control are *irrelevant* in situations of choice.

Marcus Willaschek has argued along similar lines that there are contexts where it is proper to ignore that my actions were caused by the past and the laws of nature, just like there are contexts where it is proper to ignore the possibility that an agent just dreamt what she thought had happened. Willaschek uses a court of law as example of both.¹³¹ Willaschek argues convincingly against Hawthorne and Rieber that what makes certain causes considered relevant for determining whether a decision is free is not what has been mentioned in conversation (remember that Hawthorne and Rieber argued that the distant past and the laws of nature become relevant as soon as someone brings them up), but rather the rules of relevance that guide the social practice in question.¹³² He then distinguishes between evaluative and explanatory contexts.¹³³ Evaluative contexts are those concerned with evaluating the prudence, morality, legality and so on of human decisions and actions. In these contexts, we do not consider causes of behaviour relevant unless it somehow impairs the agent's capacity to form considered practical judgements. Willaschek then offers some compatibilist arguments from Austin, Strawson and Wallace to the effect that this is how we *ought* to make

¹³⁰ Nelkin, 2009, p 571

¹³¹ Willaschek, 2009, pp 575 and 577

¹³² *ibid* p 575

¹³³ *ibid* p 577

judgements about free will in evaluative contexts.¹³⁴ My arguments do not really contradict anything Willaschek has to say, but they add something. Sometimes it is not just customary and plausible to disregard certain causes, or certain possibilities of error; in some contexts it is actually *necessary*. It is necessary when doing history to disregard the possibility that the Earth came into existence five minutes ago or that we are all brains in vats, and it is necessary when deciding what to do to disregard the possibility that all my actions were determined by something other than myself.

4.2.3. Contextualism about the Moral Words

How about words connected to moral agency and moral responsibility? Should they be given a contextualist meaning as well? After all, I have argued that one of the significant problems with the contextualism of Hawthorne and Rieber is that it cannot make sense of the moral words, whether they are context-dependent or not absurd consequences seem to follow. Gunnar Björnsson and Karl Persson have a contextualist theory of moral responsibility judgements that would avoid this problem¹³⁵ (at least if we assume that the meaning of “free will” follows that of “moral responsibility”, so that there will not be contexts where it is true that, for instance, someone was unfree while at the same time morally responsible.). According to them, judgements of moral responsibility vary depending on what we take to explain the action versus what we take to be part of the background conditions. From an everyday explanatory perspective people will seem morally responsible since their motives provide sought-for explanations of their actions, while genetics and upbringing remain in the background. In the context of a philosophy seminar where people discuss arguments for incompatibilism the opposite might be true.¹³⁶ The interesting question then becomes

¹³⁴ *ibid* p 578

¹³⁵ Björnsson and Persson, 2011

¹³⁶ *ibid* p 14

what kind of moral responsibility should guide our praxis of holding people responsible. Björnsson and Persson suggest that there may be normative reasons for regarding everyday moral responsibility rather than incompatibilist philosophy moral responsibility as the basis of responsibility-holding.¹³⁷ Willaschek on the other hand seems to lump philosophy seminars and everyday moral responsibility ascriptions together under the heading of the evaluative perspective. He then argues that it is only when seeking explanations that the entire causal chain behind an action becomes relevant, not when we seek an evaluation. A person who focuses on the distant past and the laws of nature when trying to determine whether somebody is morally responsible is thus mistaken. There are different ways to work out the details, but as soon as one comes up with some kind of norm or rule for when it is *appropriate* to make use of a compatibilist moral responsibility concept, rather than having compatibilism depend on nobody mentioning incompatibilist worries, and if one assumes that “free will” is compatibilist when “moral responsibility” is, the absurdity of Hawthorne’s and Rieber’s theories is avoided.

I have argued in Section 3.1. that talk about right, wrong and moral responsibility belongs to the practical perspective, to the context of deliberation and advice. Philosophy seminars also belong to this context insofar as we discuss moral problems, problems related to what one ought to do in various situations. Since causes of one’s action are irrelevant in this context, the moral words will have a compatibilist meaning, according to which practical freedom suffices. This is obviously not the only way people use these words in philosophy seminars, but according to my arguments that is how they ought to use them. Now what if we change perspective? What if we move from the context of deliberation and advice, to one of causal explanation and prediction? In this context, it is hard to see that the moral words would serve any function at all, except with a belief qualification. That people *believe* some things to be right or wrong, or that people have beliefs

¹³⁷ *ibid* p 22

about moral responsibility, may figure in causal explanations as well as predictions. The fact that something *is* right or wrong, or the fact that somebody *was* morally (as opposed to causally) responsible for an action would seem to be causally inert, and thus completely irrelevant from this perspective. It is not the case that the moral words take on a different meaning, for instance an incompatibilist one, when we change perspective. They simply have no role to play (except as belief components). There is thus a kind of contextualism regarding the moral words as well; in the context of deliberation and advice they have a meaning, which they lack in the context of causal explanation and prediction.^{138 139}

¹³⁸ I have argued throughout this dissertation that I will not take sides in normative-ethical or meta-ethical debates. Yet this paragraph might seem anti-realist, since many realists argue that the rightness of an action can causally explain why people believe it to be right – that the *best explanation*, in fact, of why we believe some actions to be right, is that they really are right. Still, there are also realists who believe that rightness and wrongness are causally inert. David Enoch, 2011, pp 50-83, argues that we ought to believe in moral facts because they are necessary for deliberation, not because they are necessary for explaining anything at all. Thus, the paragraph does not imply that realism is false, although it implies that many realists are mistaken on a certain point in their theory.

¹³⁹ If moral realism were true, there would be *a* theoretical perspective where the moral words have an important function to fill, namely the ontological perspective of investigating what there is. If moral realism is true, rightness, wrongness and perhaps also moral responsibility *exist* in their own right. It might still be the case that they have no role to play in the context of causal explanation and prediction.

5. Moral Responsibility

Thus far I have argued *for* practical perspective compatibilism. In the following sections I will examine a number of arguments *against* PPC (or against a group of theories where PPC is included), and show why they all fail. These arguments try to show primarily that moral responsibility, rather than moral agency in general, demands a kind of freedom different from practical freedom. However, since they all fail, all we need to establish in order to see whether moral responsibility is possible is whether the best or the true ethical theory employs this concept or not.¹⁴⁰

5.1. Desert-entailing Moral Responsibility

“Responsibility” and even “moral responsibility” can mean different things. It is therefore important to pin down which kind of responsibility we are talking about, before entering a discussion about the compatibilism of determinism and moral responsibility. There are some kinds of responsibility that obviously exist even if the universe is deterministic. It is, for instance, widely agreed that it could be *useful* to hold people responsible even if determinism is true. John Jamieson Carswell Smart is one philosopher who argues for compatibilism, because even under determinism people’s behaviour could be influenced for the better by administering praise or blame. This is all we need to justify the praxis of holding people morally responsible for

¹⁴⁰ To know whether moral agent A is morally responsible for act H, it is also necessary to know whether A is the same moral agent as the one who committed the act. I will write a little bit more about this issue in Section 2.4.1.

their actions.¹⁴¹ However, Smart is simply not discussing the same thing as incompatibilists, which he also acknowledges as he writes that most people have an idea of moral responsibility that is muddled up by confused metaphysical ideas.¹⁴² The same thing can be said about Sher's and Hobart's idea that blaming people is a way to evaluate their characters – this is not the same concept of blame as the one incompatibilists think of, when they argue that blame can never be deserved in a deterministic universe.¹⁴³ Watson discusses something he calls “aeratic responsibility” which one has for an action if the action reflects one's character, but unlike Sher and Hobart he claims explicitly that this is just one of several responsibility concepts. Although questions of whether the agent could have done otherwise or whether he in some strong sense was in control of his action are not important when we discuss aeratic responsibility, these questions may be crucial for other responsibility concepts, like that of “accountability”.¹⁴⁴ Korsgaard thinks that the act of holding somebody responsible should be guided by the virtues of love and respect. “Holding someone responsible can be insensitive or merciless; failing to hold someone responsible can be disrespectful and patronizing.”¹⁴⁵ She discusses the praxis of holding someone responsible rather than the property of being responsible. However, if we assume that one *is* morally responsible in cases where it is virtuous to *hold* one responsible, we seem once again to have a concept of moral responsibility different from the one discussed by incompatibilists.

Incompatibilists about moral responsibility and determinism typically discuss “desert-entailing moral responsibility”, and that is the concept I will discuss in this part of the dissertation. The meaning of “desert” is usually left implicit in writings on the matter, but I think some clarification is needed

¹⁴¹ Smart, 1961, p 302

¹⁴² Smart, 1961, p 305

¹⁴³ Sher, 2001, Hobart, 1934, p 18

¹⁴⁴ Watson, 2004, pp 266, 273, 279

¹⁴⁵ Korsgaard, 1996b, pp 188-212, quotation p 199

before engaging in a moral responsibility discussion. Owen McLeod claims “desert itself is a three-place relation that holds among a subject, an object, and a basis”.¹⁴⁶ Owen then argues that desert is a fairly broad notion, since we may, for instance, think that a picture deserves to be admired. However, even if we just discuss the kind of desert connected to moral responsibility, how agents can deserve praise or blame because they performed various actions, we can use the idea that desert is a three-part relation between agent, action and praise or blame. I will argue from the supposition that anyone who thinks that the fact that agents performed actions that was morally right (or subjectively right, right given their information and so on) suffices to justify praise, and the fact that they performed actions that was morally wrong (subjectively wrong and so on) suffices to justify blame, thinks that people can have “desert-entailing moral responsibility” for their actions. As soon as we bring in the need for character evaluation,¹⁴⁷ the beneficial effects of holding people responsible,¹⁴⁸ the virtues one may express by praising and blaming people and so on,¹⁴⁹ then we discuss some other kind of moral responsibility.

The responsibility I discuss for the rest of this chapter is the desert-entailing kind, although I will leave out the phrase “desert-entailing”.

Some philosophers think there is a conceptual connection between moral responsibility and punishments and rewards. I think there might be, but I am far from certain. The following, however, seems to me fairly plausible: Telling someone that we blame him or her for what he or she has done involves as an essential aspect of that activity an attempt to make the person blamed, the blamee as we may call him or her, feel guilty. Likewise telling someone that we praise him or her for his or her action involves essentially trying to

¹⁴⁶ McLeod, 2008

¹⁴⁷ Hobart, 1934, p 18

¹⁴⁸ Smart, 1961, p 302

¹⁴⁹ Korsgaard, 1996b, pp188-212

make the praisee feel proud of it. Simply thinking of people as morally responsible, or directing praise or blame to people not present in the conversation, will normally involve at least a *wish* that the agents will realise that they did wrong and feel guilty, or realise that they did right and feel proud. If we knew that terrible consequences would follow if an evil-doer were to feel guilt we might not wish that he or she did. Barring special circumstances, however, thinking of somebody as morally responsible for a good or bad act will involve wishing that he or she becomes either proud of his or her act or feels guilty for it – even if no beneficial consequences results from him or her so feeling. Now since guilt is a painful feeling and pride a pleasant one, it would follow that directly blaming someone essentially involves trying to inflict (a certain kind of emotional) pain on the blamee, and the other way around with direct praise, while holding someone responsible involves wishing (a certain kind of emotional) pain or pleasure on the responsible person. If punishment and reward are defined respectively as attempts to inflict pain or pleasure on people even if no beneficial consequences result, but simply because one thinks they deserve it, then it would follow that punishments and rewards are conceptually tied to the idea of moral responsibility. I think some such idea of punishment is what Leo Zaibert has in mind when he argues that forgiveness is to refuse to punish somebody.¹⁵⁰ I do not think that is true if one uses “punishment” the way one does in ordinary language, since in ordinary language it usually means legal sanctions like fines, prison time, inflicting bodily harm, and the non-legal equivalents like grounding or spanking a child. However, if one defines punishment as *any* attempt to inflict pain on somebody just because one thinks he or she deserves it, I think this analysis of forgiveness is at least plausible. It is also plausible that one cannot completely separate the question of when people ought to be rewarded or punished from the question of when they are praise- and blameworthy.

¹⁵⁰ Zaibert, 2009, pp 386-387

Whether this connection between praise- and blameworthiness and reward and punishment is not just plausible but also true is not something I will go into in this dissertation. Suffice to say that even if we assume its truth for the sake of argument, one can still separate the question of moral responsibility from the question of whether one ought to punish or reward behaviour in any way *beside* blaming and praising. I could think, without any inconsistency, that Gary is morally responsible for a horrible crime, and that it would be a good thing in itself if he came to realise this and feel terrible about what he has done. I could even think, in addition, that it would be a good thing if he happened to catch some painful disease or was struck by lightning, because he would deserve to have bad things happening to him – and I could still believe, without inconsistency, that neither I nor any other person would be justified in inflicting pain and suffering on him in any other way than blaming him.

Robert Kane believes that people can have ultimate moral responsibility for their actions, while he denies that this gives us the right to inflict horrible punishments on people who have committed horrible deeds. He thinks that it is always intrinsically wrong to interfere with other people's wishes.¹⁵¹ He also argues against making a strict distinction between action and omission. If A threatens to hurt B and you are in a position to do something about it, you are in a position where you are forced to choose between two bad actions: Either do nothing, which means that you by your omission facilitate A's hurting of B, or else restrain A.¹⁵² The lesser evil is to restrain A, because A was the one morally responsible for creating this whole sad situation in the first place. Since restraining A is still an evil, it must never exceed what is necessary to protect B. One ought to use as little violence as possible when taking A away from B, and if A must be locked up to protect B he should not

¹⁵¹ Kane, 1996, p 19

¹⁵² *ibid* p 21

have to suffer any unnecessary discomforts in his imprisonment.¹⁵³ This is not to deny that A has desert-entailing moral responsibility for his action. The motivation for putting up walls and bars between A and B was to protect B from violence, but this could in theory be done by imprisoning B just as well as A. The reason that it is A and not B who ought to be imprisoned is not that it would be inconvenient to lock up all potential victims instead of the criminals, or have bad consequences from doing so, but simply that it was A and not B who did wrong, so it is A who *deserves* it.¹⁵⁴ There is nothing inconsistent about Kane's views. If one has a normative ethics that employs the idea of desert-entailing moral responsibility, other considerations within the same ethical framework may forbid harsh punishments. Conversely, a person who does not believe in desert-based moral responsibility may argue on consequentialist grounds that there are great benefits to be had from an eye-for-an-eye legislation. It is important to keep these things in mind when discussing moral responsibility, to avoid confusing the issue of whether an agent is morally responsible with the issue of how we ought to treat him.

As I have already explained, I argue that practical freedom is sufficient for moral agency, but moral responsibility also requires that the true or the best moral theory is one that employs this concept. However, in the following chapters I will argue that such and such a person is "morally responsible" without adding the conditional "if we suppose that the true or the best moral theory is one which employs this concept". The conditional is simply presupposed.

¹⁵³ *ibid* p 29

¹⁵⁴ *ibid* p 22-23

5.2. Formal Arguments for Incompatibilism

5.2.1. A General Problem with these Arguments

There are a number of well-known arguments for the incompatibility of moral responsibility and determinism, or the complete impossibility of moral responsibility, in the free will debate. Considered as counter-arguments against PPC, they all fail for the same reason. They have to assume what ought to be proved, namely that the practical perspective is not what is morally relevant, and that PDR is false. The arguments I will examine in this chapter invite us to focus on features of the agent's situation that may be very interesting for a metaphysician, physicist, sociologist and so on, but are irrelevant to the deliberating agent or to a moral adviser. (This is also a problem for the thought experiments for incompatibilism and causal compatibilism that I will discuss later – they assume that one will arrive at trustworthy intuitions by taking up a theoretical perspective on agents.) In the following sections I will show that Galen Strawson's basic argument as well as van Inwagen's consequence argument both suffer from this defect. I cannot see how one could construct an argument that avoids this problem, and I place the burden of proof on anyone who says it can be done.

These arguments can be questioned on other grounds as well. The following two sections will both start by describing the argument to be discussed, pointing out some difficulties that are not specifically related to the discussion of practical freedom, and finally explain how it assumes what ought to be proved if used specifically to deny PPC.

5.2.2. Galen Strawson's Argument for the Impossibility of Moral Responsibility

Galen Strawson has tried to show that moral responsibility is impossible regardless of whether the world is deterministic or indeterministic, by his so-called basic argument. The argument goes like this:¹⁵⁵

1. When you act, you do what you do – in the situation in which you find yourself – because of the way you are.
2. If you do what you do because of the way you are, then in order to be morally responsible for what you do you must be morally responsible for the way you are.
3. You cannot be morally responsible for the way you are.¹⁵⁶
4. Therefore, you cannot be morally responsible for what you do.¹⁵⁷

He adds that it is of course possible that one is like one is because one chose at an earlier time to change one's character, but in that case, the question is just moved one step back.¹⁵⁸ The main point is that one cannot be responsible for what one does unless one is a *causa sui*, a cause of oneself – which is impossible (at least for human beings, even if we would grant that possibility to God).

There are several difficulties with this argument. The first one is determining what kind of moral responsibility Strawson actually discusses. I wrote earlier that some philosophers like to point out, to avoid confusion, that they are discussing "desert-entailing moral responsibility" rather than, for example, some consequentialist-motivated praxis of praise and blame, and I tried in the beginning of this chapter to describe how this concept dif-

¹⁵⁵ At least in one version, there are many. They do not differ substantially from each other though.

¹⁵⁶ Strawson does not write morally responsible, but URD, which is short for "ultimately truly and without qualification responsible and truly and without qualification deserving of praise or blame or punishment or reward". Strawson claims though, that URD is the concept of moral responsibility which exists in common sense morality. I will come back to this.

¹⁵⁷ Kane, ed, 2002, p 443

¹⁵⁸ *ibid* p 446

fers from other responsibility concepts. Strawson goes further and writes that he discusses necessary conditions for being “ultimately truly and without qualification responsible and truly and without qualification deserving of praise or blame or punishment or reward”.¹⁵⁹ To explain what this means Strawson refers to heaven and hell. If an agent has URD (an abbreviation that stands for Ultimately and truly without qualification Responsible and truly and without qualification Deserving of praise and blame¹⁶⁰) for her actions, then it is “perfectly intelligible” even if “morally repugnant” to propose that she should be tormented in hell for all eternity, or receive everlasting bliss in heaven.¹⁶¹ This is supposed to explain things even if one is an atheist. On reading this, one may wonder whether the conditions that need to be satisfied if somebody is to be URD are any different from the conditions that need to be satisfied if somebody is to be morally responsible at all. I for one think I have a grasp of what it means to be morally responsible for one’s actions, even in the desert-entailing sense, but I cannot really grasp the idea that everlasting torment in hell for sinners could be “morally repugnant” while simultaneously “perfectly intelligible” if the sinners in question were URD for what they did. Since I think that I can grasp some everyday concept of moral responsibility, the idea that people can deserve bad or good things, and still not understand Strawson’s URD, it seems to me like the two concepts might be different after all. Strawson though claims that they are not. He writes that nearly all human beings believe in URD,¹⁶² and that this notion is central to our common moral consciousness.¹⁶³ This is an empirical claim that I find hard to believe. To my knowledge nobody has made any surveys in which people are asked whether they believe it is intelligible to punish bad people with everlasting torment, but all experimental philosophy

¹⁵⁹ *ibid* p 442

¹⁶⁰ *ibid* p 442

¹⁶¹ *ibid* p 451

¹⁶² *ibid* p 447

¹⁶³ *ibid* p 452

done so far shows that people's intuitions on philosophically controversial questions differ. No experimental philosopher has discovered a philosophical problem where the intuitions of all or nearly all laypeople point in the same direction. It would therefore be surprising if it turned out that "nearly all"¹⁶⁴ human beings have the same intuitions regarding the intelligibility of everlasting torment for sinners, and that this is built into our common sense idea of moral responsibility.

This is the first problem with Strawson's basic argument, which might not be unsolvable, but must be solved regardless of whether his basic argument is used against compatibilists or libertarians. It is not clear at all that his URD is the same thing as the common-sense idea of moral responsibility, or as what most moral philosophers mean by moral responsibility. If his URD is something neither laypeople nor most philosophers care about, and if it requires much more freedom-wise than other kinds of moral responsibility do, then Strawson would only have proven that it is impossible to be something which few people care about being. That would fall very much short of accomplishing what he tries to accomplish in his article.

Let us lay premise 1 to one side for now (I will come back to it later), and look at premise 2. The only evidence he has for it is a couple of quotations from philosophers as well as one quotation from a layperson.¹⁶⁵ This does not show that premise 2 is something that can be taken for granted. Hobart, for one, argues against it.¹⁶⁶ But let us, for the sake of argument again, grant Strawson premiss 2. Premise 3 is still unsupported. Premise 3 rests on the assumption that one cannot be responsible for one's character simply because one endorses it, because one had some *compatibilist* control over shaping it, or even because it is part of what I am.¹⁶⁷ This assumption is not argued for in Strawson's article, it is not even made fully explicit. That the

¹⁶⁴ *ibid* p 447

¹⁶⁵ *ibid* p 454

¹⁶⁶ Hobart, 1934, p 18

¹⁶⁷ Sher, 2001

conclusion follows from the premises does not help if the premises are shaky.

Thus far I have given several reasons why Strawson has not established his conclusion. There is also one reason why Strawson's Basic Argument cannot be used for the specific task of showing that PPC is false. Although Strawson does not seem to realise this himself, his premises presuppose that a particular theoretical perspective is the right one to have when making moral judgements. This becomes clear when we look at premise 1: That I do what I do because of the way I am.

"The way I am" is an ambiguous phrase, and "because" is an ambiguous word. "Because" sometimes points to one's *reasons* for an action, but it is far from always true that these reasons refers to facts about oneself, so this cannot be the right interpretation. We could instead interpret the "because" of premise 1 in a causal sense. Let us do so, and then move on to "the way I am". There is an everyday sense of the phrase, according to which the question "what is Sofia Jeppsson like?" would be truthfully answered by citing personality traits or interests like "she's intelligent but forgetful, she's artistic and she loves dogs". But this interpretation does not make premise 1 universally true, since these kinds of interests and personality traits cannot always explain why I do what I do. There is a more scientific psychological sense according to which a true description of "the way I am" might be "introverted, agreeable" or "extroverted, disagreeable". Once again, this interpretation would not make premise 1 universally true. People sometimes act out of character and people sometimes perform actions that marks a turning point in their character development. Even if it is true that I am generally lazy, I may decide on a single occasion to put a lot of effort into something because I think it is very important, and I may also decide to put more effort into things *from now on* and implement that decision. However, if we interpret "the way I am" as describing the complete state of my brain and body, and "because" as causal, premise 1 may be universally true. It would then

mean that all actions I perform are causally explicable by detailed facts about my brain and body. However, although there is one possible causal interpretation of premise 1 that might make it true, all causal interpretations make it irrelevant for questions of moral agency and moral responsibility according to PDR. Even when an action I perform can be causally explained by one of my personality traits (for instance, I might go to a party because I am extroverted), it will often be the case that my *reason* for the action in question was something else (for instance, I want to have fun and I think parties are fun). The reason for my walking my dogs even on days when I do not feel like it is that I have a duty to care for them, including walking them, while the cause might be something like a love of animals instilled in me when I was a child. It is very difficult to conceive of a situation where the *state of my brain* could give me reason to do something. Thus, there are several causal interpretations of “because” that make premise 1 false and one that might make it true, but they are all irrelevant according to PDR. Since I use PDR as an argument for PPC, one cannot use Galen Strawson’s basic argument as a counterargument without begging the question by simply assuming that PDR is false.

Oddly enough, Strawson himself does not think that one must interpret premise 1 in the above theoretical/biological/physical way. On the contrary, he rather encourages the reader to take up a practical perspective when considering it. He writes that we should imagine what it was like when we deliberated about something:

Consider a particular action or piece of deliberation in which you engage, and consider everything about the way you are when you are to engage in it which leads you to engage in it in the way you do. I will call the particular action or piece of deliberation that you engage in “A”, and I will call everything about the way you are mentally when you engage in it that leads you to engage in it in the way you do “N”. I will use URDA(t) and URDN(t) to mean URD for A

at time t and URD for N at time t , respectively. ... When you act or deliberate, at t_1 – when A occurs, at t_1 – you do what you do, in the situation in which you find yourself, because of the way you are – because you are N, at t_1 . This is the first premise of the argument. I take it to be incontrovertible, quibbles aside, and will not defend it.^{168 169}

I will now follow Strawson's suggestion, and consider a piece of deliberation in which I have engaged. In the year 2000 I deliberated about whether to continue art school or go back to university. This is a good example, since I thought there were reasons for both alternatives, so there is no obvious necessity from previous character traits or settled preferences to the choice I eventually made. I also thought things through carefully before making my decision, so it seems like a good *prima facie* candidate for a morally responsible choice. Still, Strawson claims that nobody can be morally responsible for anything, so if his argument works, it should work equally well for this choice between university and art school. Let us start with N. If somebody asked me "what I was like mentally" when I deliberated about art school and university, I would spontaneously say that I loved painting, but I was also curious about philosophy. I was afraid that I would end up unemployed if I made the wrong choice, and I thought my prospects of getting a job would be a little better with a university degree than if I continued art school. It is obvious that these facts alone do not make a choice. I eventually chose to go back to university, but an agent could have those same mental characteristics and instead decide that the love of painting outweighs all other reasons and stick to art school. Strawson must therefore intend more than the previously mentioned mental traits to be included in N. Therefore, let us add that I found my belief about the high unemployment risk for artists to outweigh

¹⁶⁸ Kane, ed, 2002, pp 444-445

¹⁶⁹ One small oddity about this quote is that cause and effect are described as simultaneous. Both A and N are at t_1 , although one would have expected N to come first. However, I will not dwell on this.

my love of painting, and that I did not suffer from weakness of will, but did what I judged to be best. Now it is easy to see how N leads to a certain choice on my part, to go back to university, but it is hard to see how Strawson would get a responsibility-threatening regress started. The regress demands that I am not responsible for N. However, on this account, it seems like I am responsible, at least for part of N. Even if I am not responsible for being curious about philosophy or loving to paint, I seem to be responsible for weighing the reasons the way I do. However, Strawson denies that one can be responsible for any part of N, because that would require that one intentionally brought it about that one became N.¹⁷⁰ I did, in fact, intentionally bring it about that I weighed the reasons pro and con art school and university in a certain way. Strawson argues that this intentional bringing about must have been caused by the way I was mentally at an earlier time, and he calls the way I was mentally then M.¹⁷¹ Yet, it is simply not true that I can look back to some earlier point in time and find mental characteristics there that lead me to weigh the reasons in exactly the way I did when choosing between art school and university, and we have no reason to believe that there are psychological laws according to which my previous mental characteristics lead to all the details of N.

The only way to get the regress started is to skip the practical perspective and look at my psyche from outside, as if it were a piece of machinery – to regard the weighing of reasons as just another event caused by previous events. As long as I keep the practical perspective of choice and consider a particular piece of deliberation in which I engage, and everything about myself that makes me engage in it the way I do, I will only discover how I weigh various reasons pro and con, not some responsibility-threatening regress of determination and/or randomness. Although Strawson seems to encourage the reader to take up a practical perspective when reading his first

¹⁷⁰ Ibid p 446

¹⁷¹ ibid pp 446-447

premise, this same perspective destroys his argument. Premise 1 must be interpreted in a particular way, as saying that the state of my brain and body causes bodily movements, in order for the argument to be valid and freedom-undermining at the same time.

In some parts of the text Strawson seems half-aware of the fact that there are different ways to view an agent depending on which perspective one takes up. In Chapter 6, called “the irrelevance view and the agent-self” he writes about an agent choosing and how from his own first-person perspective he must *make* the choice *himself*.¹⁷² He discusses the idea that desires in themselves does not make a decision until the agent has decided whether to take them as reasons for action or not. Then, Strawson steps back again to a theoretical perspective and views the agent’s mentality as a kind of machinery. From this theoretical perspective the agent can be reduced to just further mental mechanisms whose mode of operation decides the outcome. Strawson simply assumes that this is the morally relevant perspective, and since from this perspective there are only long causal chains and/or quantum randomness and no crucial difference between agents and other things in the universe, moral responsibility does not exist. He must simply *assume* that this detached physical perspective is the one that provides us with morally relevant information, or else the argument does not get off the ground. He must, in short, simply assume that PDR is false and that many things irrelevant to a deliberating and choosing agent are highly relevant when deciding whether somebody is morally responsible. I, on the other hand, have argued for the relevance of the practical perspective and the irrelevance of the theoretical one; I have not simply assumed it. Therefore the basic argument cannot be used as a counterargument to PPC, since it would be assuming what ought to be proved.

¹⁷² *ibid* pp 455-458

5.2.3. The Consequence Argument and Frankfurt Responses

Peter van Inwagen's famous consequence argument attempts to show that if determinism is true, nobody could ever have done otherwise. In one version, van Inwagen's argument is about a judge, J, who at time *t* is in a position to save a man from being executed by lifting his hand. However, J kept his hand down. It is stipulated that J is physically healthy, there is nothing wrong with his arm or hand, and neither is there anything abnormal about his psychology.¹⁷³ He basically fulfils what a compatibilist would ask for when it comes to moral responsibility for his actions. Now if determinism is true, van Inwagen argues, then "J did not lift his hand at *t*" is implied by the conjunction of a proposition describing the state of the universe in the remote past and a proposition describing all laws of nature. For J to have the ability to lift his hand, it must be the case that J has the ability to do something such that if he did it either the laws of nature or the remote past would have been different. However, J does not have the ability to do something such that if he did it the laws or the past would have been different. Therefore, J has no choice but to let his hand lie down.¹⁷⁴ (It should be noted that van Inwagen does not attempt to show that there is a *special sense* in which J did not have the ability to lift his hand. He wants to argue that J completely lacks the ability to lift his hand. He claims in a forthcoming paper that "compatibilists and incompatibilists mean the same thing by 'able'. And what do both compatibilists and incompatibilists mean by 'able'? Just this: what it means in English, what the word means."¹⁷⁵)

¹⁷³ van Inwagen, 1983, p 69

¹⁷⁴ Ibid, p 70

¹⁷⁵ van Inwagen, forthcoming in *Ethics*, p 10. Here he also claims that it was a mistake to use the word "could" in the original version of the consequence argument, since "could" is ambiguous. Therefore, I consistently use the phrase "were able to" instead, which van Inwagen himself prefers.

The consequence argument, if it works, threatens not only compatibilism about moral responsibility and determinism, but also compatibilism about moral agency and determinism. It has often been supposed that normative-ethical theories that do not include the notion of moral responsibility are more naturally compatibilist. The main contestant to theories invoking the notion of moral responsibility is utilitarianism. However, utilitarianism is not any less threatened by the consequence argument than theories employing moral responsibility. It may seem as if utilitarianism and compatibilism are natural companions, since actions can have good or bad consequences even if determinism is true. But according to classic act utilitarianism, it is *not* the case that an action is wrong just because the consequences were bad – if the action was unavoidable, it was morally right regardless of its consequences. Moral wrongdoing only occurs when there was at least one other option available that would have had *better* consequences than the one performed. Thus, moral agency requires the ability to do otherwise according to utilitarianism.¹⁷⁶ If van Inwagen’s argument holds and judge J did not have the ability to lift his hand, then keeping his hand down was trivially the right thing to do according to utilitarianism, since he had no other option with better consequences. However, many utilitarians distinguish between blameworthy and wrong as well as praiseworthy and right. Perhaps somebody is tempted to say that judge J committed an act of blameworthy wrongdoing – even if his action was right, simply because it was unavoidable, he should be blamed for failing to save the man’s life. However, if we take “blameworthy” to mean that it is *right* to blame the person in question, then “blameworthy” according to act utilitarianism just means that blaming the agent will have better consequences than not blaming him, and the other way around. Let van Inwagen’s example be about blaming instead of hand-lifting. The exam-

¹⁷⁶ Remember that my definition of moral agency said that although the agent can be an angel or a devil, he or she must exist in a universe where both good and bad refers to something or can be used correctly in some situations. If there is no possibility to do otherwise in the universe, and thus no possibility of wrongdoing, there is no moral agency in my definition.

ple could just as well be used to show that a blamer could not have done anything but blame. If they could not have refrained from blaming, then it is trivially true that blaming had the best consequences of all available options, and was thus right. Every time somebody blames somebody else it is trivially right to do so, and "blameworthy" reduces to "receives blame" (and the same reasoning holds for praise). A utilitarian cannot claim that it was wrong to blame somebody just because *the blamee* could not have done otherwise, but he *must* claim that it is only wrong to blame somebody in those cases in which the *blamer* could have done otherwise. If nobody ever could do otherwise, this includes blamers, who are therefore always doing the right thing when they blame somebody. Therefore, if the consequence argument is a threat to morality at all, it is as much of a threat towards act utilitarians as towards believers in moral responsibility.

When it comes to rule utilitarianism, I suspect that the problem remains. If one could not have done anything other than adopting a certain rule, then this rule is trivially the one with the best consequences of all the rules one could have chosen. Perhaps there is some particular version of rule utilitarianism that escapes this problem, and perhaps there are specific deontological or virtue-oriented theories for which the consequence argument becomes irrelevant, but I will not investigate the matter further. I just think it is important to point out that one cannot escape the force of this classic incompatibilist argument by simply declaring oneself a utilitarian. If anything, the consequence argument is *more* of a threat towards utilitarian wrongdoing than to moral responsibility. Many people believe that Harry Frankfurt (to be discussed below) has provided a good counterargument towards van Inwagen by showing that the ability to do otherwise is unnecessary for moral responsibility, but this argument will be irrelevant for the utilitarian. Since the idea of alternative possibilities is part of the very definition of wrongdoing on a utilitarian account, no argument could prove that alternative possibilities were not necessary for utilitarian wrongdoing. There is thus at least

arguably a compatibilist argument available for the moral-responsibility-theorist that a utilitarian cannot use.¹⁷⁷

This is why the free will debate is as important to utilitarians as it is to deontologists.

The above version of the consequence argument is directly about the ability to do otherwise and only indirectly about moral responsibility. A popular strategy among compatibilist philosophers is to invoke so-called Frankfurt examples to show that agents *can* be morally responsible for unavoidable actions. If they can, the so-called principle of alternative possibilities, or PAP for short, according to which an agent can only be morally responsible for an action in case he could have done otherwise, is false.

A classic Frankfurt example looks like this: Black wants Jones to kill Smith. Black hopes that Jones will do this on his own, but to ensure that Smith ends up dead he installs a device in Jones' head without Jones' knowledge. The device enables Black to take control over Jones' thoughts and actions if need be. If Jones is about to abstain from murder, Black will step in and make him murder Smith anyway. As it happens Jones murders Smith of his own accord and Black remains idle. Many philosophers have taken this to show that PAP is false. Jones could not but murder Smith, and yet he is supposedly responsible for it. However, everybody has not been convinced.

Independently of each other, David Widerker and Robert Kane discovered the so-called dilemma problem for Frankfurt examples. The dilemma is between stipulating that the example is set in a deterministic universe and stipulating that it is set in an indeterministic one; presumably, it must be one or the other. Now suppose first that the whole story takes place in a deterministic universe. In this case everybody with incompatibilist intuitions will

¹⁷⁷ See also Haji, 1999, for the view that determinism precludes right- and wrongdoing, though not moral responsibility, since the latter is saved by Frankfurt.

feel that Jones, simply because his actions were determined by the past and the laws of nature, cannot be morally responsible for what he does, Black or no Black. If we instead suppose that the universe where they live is indeterministic and Jones is a libertarian agent who basically satisfies libertarian demands for moral responsibility, then it is difficult to see how Black could manage to rob him of all alternative possibilities. Black may be very good at predicting Jones behaviour, but it is hard to see how infallible prediction could be possible if the world is indeterministic. If Black's predictions are not quite infallible, it *could* happen that Jones decides to abstain from murder while Black mistakenly thinks he will murder Smith of his own accord and remains idle; it follows that Jones does have alternative possibilities.¹⁷⁸

In response to this dilemma, Alfred Mele, Pereboom and others have tried to construct examples that take place in an indeterministic universe but where, thanks to some extremely complicated machinery inserted in Jones' head or some very elaborate psychology on his part, all other possibilities are still ruled out. I think the problem with these responses is that intuitions reach some kind of melt-down point trying to deal with scenarios of this complexity. This is true not only about my own intuitions; Pereboom has said that people often react this way to his Frankfurt case "tax evasion", although he does not consider the scenario very complex.

Tax Evasion: 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 get caught, and that if he is, he can convincingly plead ignorance. Suppose he has a very powerful but not always overriding desire to advance his self-interest no matter what the cost to others, and no matter whether advancing his self-interest involves illegal activity. Furthermore, he is a libertarian free agent. Crucially, his psychology is such that the only way that in this situation he could fail to choose to evade taxes is for moral reasons. (As I use the phrase here, "failing

¹⁷⁸ Widerker, 1995, and Kane, 1985 p 51

to choose to evade taxes” will encompass both 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 fact, it is causally necessary for his failing to choose to evade taxes in this situation that a moral reason occur to him with a certain force. A moral reason can occur to him with that force either involuntarily or as a result of his voluntary activity (e.g., by his willing to consider it, or by his seeking out a vivid presentation of such a reason). However, a moral reason occurring to him with such force is not causally sufficient for his failing to choose to evade taxes. If a moral reason were to occur to him with that force, Joe could, with his libertarian free will, either choose to act on it or refrain from doing so (without the intervener’s device in place). But to ensure that he chooses to evade taxes, a neuroscientist now implants a device which, were it to sense a moral reason occurring with the specified force, would electronically stimulate his brain so that he chooses to evade taxes. In actual fact, no moral reason occurs to him with such force, and he chooses to evade taxes while the device remains idle.¹⁷⁹

In a way Pereboom is right; the scenario is not very complex; Joe is just some guy who decides to cheat on taxes. I think the problem is that when Joe’s psychology is given a detailed *causal* description, as it must have if the example is going to work as a Frankfurt case, then Joe starts to look like a complex machine rather than an agent. There is Joe’s moral interest, his self-interest, the psychological mechanism that makes him unable to abstain from the crime unless a moral reason occurs to him with a certain force, and so on. This impression remains despite libertarian free will being inserted like one of the cogs in the machinery. At least I think this is the reason why my own intuitions start to slide off into an incompatibilist direction while reading texts like that of tax evasion, despite the fact that I am a compatibilist to start with. Frankfurt examples need to be fairly simple and straightforward in

¹⁷⁹ Pereboom, 2001, pp 18-19 of the chapter “Alternative Possibilities and Causal Histories”

order to intuit strongly regarding the agent's moral responsibility in the audience, but the examples need to be complex for the indeterministic horn of the dilemma problem to be solved.

John Martin Fischer, on the other hand, have tried to show that the assumption that the universe is deterministic and that Black therefore is able to predict Jones' behaviour with certainty need not be question-begging against incompatibilists.¹⁸⁰ We need not presuppose that Jones can be morally responsible despite living in a deterministic universe. We can remain agnostic on that point. We can even remain agnostic as to whether determinism in itself rules out the ability to do otherwise. We can still see that Black and determinism taken together rule out all alternative possibilities for Jones. By reflecting on the example, we can then see that the fact that Jones lacked alternative possibilities does not affect his moral responsibility; *if* determinism somehow robs him of moral responsibility it must be for some *other* reason.¹⁸¹ At least we will reach that conclusion if we agree with Fischer that Jones does not have alternative possibilities in the morally relevant sense. If murdering-Smith-on-his-own and murdering-Smith-under-Black's-control count as two different options, Jones will always have another alternative. Michael Otsuka has argued that one can only be morally responsible for an action if there was an alternative option for which the agent would have been blameless. This will always be true of Jones, regardless of how one fills in the details of the thought experiment, since he will be blameless if Black takes control over him.¹⁸² My own Frankfurt case intuitions tend to go in the same direction as Otsuka's, although the majority of the philosophers engaged in the debate seem to agree with Fischer that it cannot be morally relevant whether Jones had the option of performing the same action on his own or controlled.

in the ebrary electronic book

¹⁸⁰ Fischer, 2010

¹⁸¹ Fischer, 2010, p 14

¹⁸² Otsuka, 1998.

Let us, however, suppose for the sake of argument that Fischer has shown that the dilemma problem can be solved on the deterministic horn. Let us suppose that Black robs Jones of all alternative possibilities while Jones remains morally responsible. It follows that PAP is false. How could this be used as an argument for compatibilism? After all, as Frankfurt himself noted,¹⁸³ the reason we tend to hold Jones responsible despite him lacking alternative possibilities is that Black does not intervene. Crucially unlike the past and the laws of nature in a deterministic universe, Black waits passively behind the scene, only to interfere if Jones is about to make the wrong choice. Since there are many cases where we do accept “he could not have done otherwise” as a valid excuse, we might conclude from the story of Jones and Black that an agent *do* need alternative possibilities in order to be morally responsible, *except* in those cases where the alternative possibilities are removed by something that does not affect the actual chain of events.¹⁸⁴ This might sound terribly ad hoc, but it is hardly more ad hoc than Frankfurt’s own suggestion. Frankfurt first hypothesises that PAP is replaced by a principle according to which one cannot be morally responsible for an action if one did it *because* one could not do otherwise. This principle would be consistent with Jones’ moral responsibility, while explaining why “he could not do otherwise” often is a valid excuse. However, this principle could, as Frankfurt points out, be given an incompatibilist interpretation.¹⁸⁵ One could argue that an agent in a deterministic universe does what she does because determinism made anything but this action physically impossible, or, in other words, because she could not do otherwise. Frankfurt, however, wants to argue for compatibilism. Instead, he therefore replaces PAP with PAP*, according to which an agent cannot be morally responsible for an action if he

¹⁸³ Frankfurt, 1969, pp 836-837

¹⁸⁴ See also Glatz, 2007, for a similar suggestion

¹⁸⁵ Frankfurt, 1969, p 838

did it *only* because he could not have done otherwise.¹⁸⁶ This principle, he contends, does not threaten moral responsibility under determinism.¹⁸⁷ Presumably normal determined agents do not do what they do *only* because they could not do otherwise; it is *also* the case that they do what they want. I think this claim about determined agents is problematic; it is analogous to saying that the car does not roll forward *only* because I step on the pedal; it is *also* the case that the wheels turn around. And this principle is not really supported by the story of Black and Jones. It would be misleading to say about Jones that he does not kill Smith *only* because he could not do otherwise; it is *also* the case that he wants to, when his own wish is the *sole* cause. The story of Black and Jones intuitively supports Frankfurt's *first* suggested replacement for PAP, according to which one cannot be morally responsible for an action if one did it because one could not do otherwise, but this is not the principle that gets us compatibilism.

Finally, even if Frankfurt would have successfully shown that PAP* should replace PAP, and PAP* does not support incompatibilism, one can argue, with Pereboom, for source incompatibilism. According to this position determinism is a threat because it means that the source of our actions is something outside us over which we have no control, namely the distant past and the laws of nature.¹⁸⁸ The consequence argument can be used to support source incompatibilism as well. It tells us that J lacks the ability to lift his hand *because* he does not have the ability to do something such that if he did it the past or the laws would have been different. The crux of the argument, therefore, is that J's action is the consequence of things over which he has no control. This becomes even clearer in van Inwagen's so-called sketchy version of the argument:

¹⁸⁶ *ibid*, p 838

¹⁸⁷ *ibid*, p 839

¹⁸⁸ Pereboom, 2001, from the chapter "Introduction to Hard Incompatibilism"

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.¹⁸⁹

That the crux of the argument is how actions are caused by factors over which we have no control is also clear from the following so-called direct argument van Inwagen invented to support incompatibilism about moral responsibility and determinism. In this argument, he uses NP as an abbreviation for “no human being or group of human beings is even partially morally responsible for P”. He uses two inference rules in the argument, A and B:

A: If P is necessary it follows that NP

B: If NP, and $N(P \rightarrow Q)$, it follows that NQ.

The argument looks like this:

1. If determinism is true, and we substitute S for a description of some state of the world in the remote past, before there were any humans around, L for a complete description of all laws of nature, and T for any true sentence describing something somebody has done, then it is necessary that S & L \rightarrow T. (Premise)

2. We can deduce by elementary modal and sentential logic, that it is also necessary that $(S \rightarrow (L \rightarrow T))$.

3. N $(S \rightarrow (L \rightarrow T))$. (From 2 by A)

4. N S. (Premise)

5. N $(L \rightarrow T)$. (From 3 and 4 by B).

6. N L. (Premise)

Conclusion: N T. (From 5 and 6 by B)¹⁹⁰

This argument omits all mention of alternative possibilities and having no choice about one’s action. However, anything mentioned in T will be the

¹⁸⁹ van Inwagen, 1983, p 56

consequence of factors nobody can be morally responsible for, the distant past and the laws of nature, and therefore, according to the argument, nobody can be morally responsible for T either.

I will now respond to the consequence argument and the direct argument.

As for the consequence argument, I will not deny the importance of having alternatives or having a choice. I have, after all, argued at length for a conception of freedom according to which we are free when we must choose our actions. However, the consequence argument can only show that J lacks freedom or choice in a certain theoretical sense. John Perry has distinguished between two different meanings of “ability” – agents have “the strong ability” to perform an action if they could do it given the actual past and the laws of nature, while they have “the weak ability” to perform an action if something like the conditional analysis (perhaps with added conditions, such as there being no psychological barriers against choosing the action) is true.¹⁹¹ Now J may lack the strong ability to lift his hand, but what strong abilities one has are irrelevant when deliberating about what to do (at least as long as we lack omniscience and do not know which actions we are determined to perform), and therefore, according to PDR, irrelevant when judging whether somebody is morally responsible or not. I thus disagree with van Inwagen about ability; it is clearly a word that can be used in different ways, as has been shown by for instance Perry’s writings about strong and weak abilities and subsequent philosophical discussions. If van Inwagen wants to maintain that “ability” is completely non-ambiguous and only means one thing – “what it means in English, what the word means”¹⁹² – he must maintain that Perry as well as everybody who has taken his writings about weak and strong abilities seriously and discussed them are linguistically confused and talk nonsense. This is clearly not the case. So van Inwagen has merely

¹⁹⁰ van Inwagen, 1980, p 32

¹⁹¹ Campbell, ed, 2004, p 231 and pp 241-245

shown that there is *a sense* in which J lacked the ability to lift his hand, and I deny that this is the morally relevant sense.

What is relevant for a deliberator are weak abilities, both the ones we actually have and the ones we merely think that we have.¹⁹³ To see how weak abilities that we only think that we have can be morally relevant, let us go back to Jones in the Frankfurt example. He lacked not only the strong but also the weak ability to abstain from murder, but he had no way of knowing about the latter. As long as he thinks he has the weak ability either to murder Smith or not, he must choose what to do, and has at least minimal practical freedom. He is not open to all kinds of moral judgements according to all ethical theories, but that he deserves blame is one of the judgements we may plausibly make despite his false beliefs about the efficaciousness of his will, since there were two different options he could rationally deliberate about. However, what is relevant for a moral adviser who knows about such circumstances as Black and devices in the head are the weak abilities the agent actually has. Jones could not be advised to abstain from murder by someone who knew about Black. J, on the other hand, could be advised to lift his hand. J could both rationally deliberate about lifting his hand and be advised to do so. Since there are no obstacles to his practical freedom, he is open to every kind of moral judgement. I have already argued that third-person judgements must be such that they could hypothetically function as advice and therefore have the same rules of relevance. Since strong abilities are not relevant for advisers they are not relevant when making third-person judgements either. van Inwagen may have proven that there is *a sense* in which nobody could ever have done otherwise if determinism is true, but it is not the morally relevant sense.

¹⁹² van Inwagen, forthcoming, p 10

¹⁹³ Weak abilities are also what interest scientific disciplines such as medicine, psychology etc. What strong ability a person has to perform a certain action seems to be of interest solely to free will philosophers.

When it comes to van Inwagen's direct argument, I must deny inference rule B. I have argued that moral judgements should be made from a practical perspective, according to which the causes of one's behaviour, when they cannot be figured as reasons pro or con choosing an action, are morally irrelevant. For the same reason entailment relations between descriptions of the distant past and the laws of nature and some action I perform are morally irrelevant. If we imagine that P stands for a description of some state of the world in the remote past and all laws of nature, while Q is some action of mine, it is true that I am not morally responsible for P, nor for the fact that P entails Q. However, since the fact that $P \rightarrow Q$ should be considered irrelevant when making moral judgements, it does not follow that I am not morally responsible for Q. If we let P stand for something other than a description of the remote past and the laws of nature, it may *seem* as if the inference rule obviously holds. Imagine, for instance, that I was kidnapped and placed in a cargo car on a train on its way to Gothenburg. Let us suppose that facts about the train, the cargo car (I try as hard as I can to alert the staff on the train, but the cargo car walls are thick enough to insulate any sound I make) and the train route imply that I will be in Gothenburg at time T. I am not responsible for being kidnapped, for the design of the cargo car or the route of the train, and neither am I responsible for being in Gothenburg at time T. However, the latter lack of moral responsibility is not transferred from the former. Rather, my lack of responsibility for being in Gothenburg at T depends on my lack of practical freedom in this situation. My will is inefficacious regarding being in Gothenburg at T, and I know this. Thus, according to PDR, van Inwagen's inference rule is not valid, but PPC can still explain ordinary judgements about people's lack of moral responsibility.

Since van Inwagen wrote this article some philosophers have devised counter examples to B (or rather to rules implied by B as originally stated),

which have prompted van Inwagen to modify it slightly.¹⁹⁴ He now prefers a formulation in terms of having access to regions of logical space. Perhaps further counter examples and further small modifications of the inference rule will appear in the future. Regardless of how the details of B are worked out, as long as it says that lack of responsibility for the laws and the past can somehow transfer to lack of responsibility for actions, I will deny it for the very same reason as given above: When making moral judgements, the fact that $P \rightarrow Q$ when P stands for some state of the world in the distant past and Q an action, should be considered irrelevant.

5.3. Thought experiments for incompatibilism

In the previous section I already discussed a famous thought experiment often used by compatibilists, namely the Frankfurt example. Incompatibilists also use thought experiments frequently, in addition to more formal arguments such as the basic argument and the consequence argument. These thought experiments usually have the following form: “X in situation S is clearly not morally responsible for his actions. Since normal people under determinism are analogous to X in S, normal people are not morally responsible either”. Like the formal arguments, however, these thought experiments tend simply to presuppose that some detached theoretical perspective is the relevant one when judging whether or not people are morally responsible. They assume that by focusing on the causes of an action rather than taking up the agent’s perspective and considering his reasons for it, we will arrive at trustworthy moral intuitions.

¹⁹⁴ See Kane, 2002, ed, pp 158-167 for a discussion of the corresponding B rule for an argument for the incompatibility of free will and determinism, and how exactly B should be reformulated to avoid counter-examples.

5.3.1. Hardline and Softline Replies

Michael McKenna thinks that compatibilists too often focuses on pointing out faults in incompatibilists' thought experiments that make the situations there described disanalogous to normal situations under determinism. He calls this strategy "soft-line reply," and he thinks it has serious limits. Sometimes this kind of response is appropriate, but far from always. Even if one can point out problems with particular formulations of thought experiments, there is nothing to stop incompatibilists from inventing new ones that more closely mimic the conditions under determinism. McKenna's suggestion is to focus more on "hardline replies", to argue that the main characters in those stories *are* morally responsible for what they did.¹⁹⁵ One can do so by simply focusing more on the relevant agential properties of the characters. A hardline reply does not have to convert incompatibilists – but if it seems plausible to people who already have a compatibilist leaning, it means that incompatibilists cannot use the thought-example to convert people either, and that is what those thought examples were intended to do in the first place.¹⁹⁶

That one can come up with hardline replies to existing examples does not prove that nobody will ever invent a thought experiment where a plausible hardline reply is impossible. Still, it seems like the burden of proof lies on the one who claims that such a thought example *can* be constructed.

My theory implies a certain kind of hardline reply to common thought experiments that elicit incompatibilist intuitions by making the audience focus on the causes of the agent's behaviour. I claim that causes of behaviour that neither destroy the agent's capacity for deliberation nor can figure as reasons for action are morally irrelevant, and thus I must claim that all thought experiments designed to make the audience focus on such causes make them focus on factors that are morally irrelevant. In a way, I am even

¹⁹⁵ McKenna, 2008, pp 143-144

more hardline than McKenna. McKenna writes that if compatibilists, by focusing on more compatibilist-friendly features of the scenario, can keep their intuitions intact, the incompatibilists have failed to accomplish what they tried to do with their thought experiments, which is to convert people. I say that everyone (not just people with a natural compatibilist leaning) *should* focus on only those features of the example that are relevant to deliberation and advice. Still, I will not settle for merely pointing out in response to every thought experiment that according to PDR one ought to disregard causes that are irrelevant for a deliberator or adviser when judging whether someone is morally responsible. I will also try to show that PDR does not force me to accept any wildly counter-intuitive implications, that there is nothing very strange about judging the agents in the discussed thought experiments morally responsible.

5.3.3. A Softline Reply for Vehicle Scenarios

One kind of thought experiment features agents riding different vehicles, thinking they are free when they are not. Joel Feinberg imagines life as a train-ride, where the track over and over again branches out, and the agent driving the train can decide which branch to follow. In this scenario, the agent is free to go wherever he wants to. However, if all branches are closed so that he is forced to drive the train straight ahead, he has no liberty, even if he wants to go straight ahead.¹⁹⁷ Fischer has come up with a similar example that is supposed to show that an agent under determinism has what he calls guidance control, though not regulative control.¹⁹⁸ He imagines riding a car with a faulty steering wheel. It works as it should when he turns left, but if he would try to turn right or go straight ahead the car would somehow go to the left anyway. This means that Fischer has guidance control over the car

¹⁹⁶ *ibid* pp 147-148

¹⁹⁷ Feinberg, 1980, pp 30-44

¹⁹⁸ Fischer, 1994, p 133

when he turns left, since it is his turning of the wheel that causes the car to turn, but not regulative control, since he could not have turned any other way. There are many other vehicle scenarios in the literature, where the person riding the vehicle thinks he fully controls it but is mistaken. I think that all these examples fail to say anything about agents under determinism, no matter which perspective one takes. Fischer's and Feinburg's men have this in common: They could not turn their vehicles any other way even if they tried to. Determinism is not like that. If I would have tried to watch TV instead of write on my computer, it is not as though determinism (or the laws or what-have-you) would have stopped me.

If one feels some incompatibilist pull from considering vehicle thought examples, where the driver would not be able to drive any other way had he tried to, I think this pull may very well depend on those features of the example that are disanalogous to normal situations under determinism.¹⁹⁹ At least for my own part I would very much like to have a car that goes left when I want to go left and goes right when I want to go right. If I learnt after a long drive that there had actually been some weird fault with the steering wheel, so that I only had guidance control, I would become terrified and think that I had been driving through traffic with no real control over my car at all and survived by sheer luck. Since determinism does not mean that we are in this situation, I will leave all vehicles aside.

¹⁹⁹ Fischer's example is not intended to support incompatibilism but "semi-compatibilism", according to which determinism precludes the ability to do otherwise but not moral responsibility. Still, I have encountered students who felt that if the control we can have over our actions under determinism is analogous to the control the driver has over the car with the faulty steering wheel, we have no control at all and could not be responsible for what we do.

5.3.4. A Hardline Reply for Walden Two

Behaviourist psychologist Burrhus Frederic Skinner wrote a utopian story called *Walden Two*,²⁰⁰ which exemplifies what incompatibilist philosopher Robert Kane calls CNC (covert and non-constraining) control.²⁰¹ In the story, a society is based on behaviourist principles and arranged in such a way that its citizens are conditioned from birth to be kind, harmonious and happy. Kane claims that one problem for all compatibilists is that they cannot explain what is wrong freedom-wise with the people in Skinner's story. Now Kane agrees that "freedom" can mean different things, and that being free from compulsion, prison and so on is to be preferred even if one could not be free in a deeper sense. So he agrees that the Walden Two citizens are free in certain respects, but he thinks it obvious that they do not have the kind of freedom required for moral responsibility and thinks it is a problem for compatibilists to explain why.

The people in Walden Two, according to Kane, satisfy anything a compatibilist could ask for. I agree on this point. The inhabitants are described as benevolent, intelligent, reflective people who can give good reasons for their specific actions as well as their general way of life, and if some particular compatibilists think there is some little thing lacking it would probably be easy enough to build it in.²⁰² ²⁰³ Walden Two is a village founded by people

²⁰⁰ Skinner, 1948

²⁰¹ Kane, 1998, p 64

²⁰² The phrase "disgustingly perfect" also comes to mind when reading the book. After a while, it is easy to be annoyed with the Walden Two citizens simply because Skinner goes to great lengths to explain how happy and perfect they are. Perhaps some people feel an incompatibilist intuition when they read this book, simply because the people in Walden Two do not seem like real human beings. It is pretty clear, however, that Skinner did not intend this impression. The inhabitants of Walden Two are *supposed* to be rational, intelligent, happy, and so on.

If something were psychologically amiss with these people, compatibilists and incompatibilists alike could agree that they are not quite free. If we are going to use Walden Two as a thought experiment about people who have compatibilist but not libertarian free will, we must therefore suppose that they are what Skinner intended them to be, rather than what they sometimes seem like due to Skinner's modest talent as a writer. This may be difficult if one has

who wanted to use behaviourist psychological science to create as happy a society as possible. The idea is that behaviours that are followed by pleasurable consequences are reinforced and become more frequent, while behaviours that prove fruitless are abandoned, and this is true of humans as well as of simpler animals. Society is therefore constructed in a way such that good behaviours are always rewarded, bad behaviour never. This is the basic idea. There is no censorship in Walden Two, so everyone knows about the psychological theories behind their society and how they themselves are affected by living there. They also know all about the world outside Walden Two (they trade with regular American communities, some of the youth go to university in regular cities, and so on). Anyone who takes a special interest in psychology and planning might get a psychological education and become a planner of Walden Two, so there is neither dictatorship nor class differences either.

Skinner's story is set in a decidedly deterministic universe, but the deterministic laws are psychological rather than physical. Everyone, whether a Walden Two citizen or not, always acts as they do because of the environmental influences and past experiences they have been subjected to. Outside Walden Two, people may act irrationally, self-destructively, evil, and be torn by conflicting impulses, which means that they might be difficult to predict. In Walden Two on the other hand people are always rational, do what makes them and others happy, and do not experience the same kind of conflicts that people who suffer from bad upbringings do. Skinner describes these people as lacking the ability to do otherwise, but what he means is merely that the citizens of Walden 2 never make cruel, self-destructive or otherwise bad

actually read the book, but should not be too difficult for people who merely read my summary of it.

²⁰³ The founder of the society, Frazer, does not come across as a very nice person, but this is explained in the book by his upbringing in normal society and its effects, which the Walden Two environment did not manage to eliminate. It is important to remember that he is not the ruler of the place, only the one who first came up with the idea. If there had been an unfair

choices. Since they do not have unconsidered options that would have been better according to their own values than the actions they choose to perform, the citizens of Walden Two actually have maximal practical freedom according to PPC, while normal people fall short of this. Kane writes about philosophers who “bite the bullet” and claim that the Walden Two citizens *are* perfectly free and responsible²⁰⁴ – and I would be the biggest bullet-biter of them all, since I claim that the Walden Two citizens are even freer than regular people. But I object. I do not think there is even a bullet to bite here. One should not be fooled by Skinner’s provocative language, where he mocks philosophers and claim that there is no such thing as free will. If one focuses on the facts of the scenario rather than Skinner’s language, the bullet evaporates. We see this more clearly if we first try to imagine a single family rather than an entire society.

Suppose two parents successfully raise their child according to the basic Walden Two principles. They never have unrealistic expectations for their child, never demand anything from him that he cannot live up to. But all efforts to be good are praised and rewarded. They never use any harsh punishments, but simply show the child that being bad does not pay off. They expose him only gradually to life’s various hardships, so that he can grow gradually stronger morally and psychologically rather than being beaten down. They encourage him to learn about the world as he grows up, since they know that he has all the necessary qualities for making wise choices and does not have to be held in ignorance about anything. Because of this (unrealistically, of course) perfect upbringing the child grows up to be both of a happy disposition and very virtuous, and may thus seem pretty predictable at times, since you can always count on him doing the right thing rather than being cruel, self-destructive or irrational. This person would still have practical freedom. That somebody had a perfect childhood is clearly irrelevant

ruler of the place, that would have afforded the compatibilist a very simple reply to Kane of what is wrong with Walden Two.

when deliberating about what to do,²⁰⁵ and thus irrelevant to practical freedom. My theory implies that he is still a moral agent and morally responsible for what he does, but I cannot see this as bullet-biting. Just think about it; would it not be terribly *strange* to, contrary to my theory, accuse the parents of robbing the child of responsibility-grounding free will? To say with indignation “But you *never* gave him conflicting messages, you *never* yelled at him for something that was not his fault, you *never* expected things from him that he could not have lived up to – you ought to have done this at least *sometimes!*” Would it not be strange, if one met him as an adult who was really nice, intelligent, caring and everything else one might wish for, to say that he could not be morally responsible since his up-bringing was too perfect? I grant that it is not implausible to claim that one should give *more* praise to a person with a more regular up-bringing who performed a good deed, than to this man. A good deed is at least arguably more praiseworthy and deserving of greater rewards if performed by someone who had to put more of an effort into doing good, or who had to put more of an effort into developing a kind and benevolent character in the first place. We might therefore think that the perfect man with his perfect parents merely deserves a little praise for his goodness, since being good is so easy for him, while the man from a harsh background who becomes good despite bad odds deserves more.²⁰⁶ However, to say that the man in my example cannot be regarded as *morally responsible* for what he does, because he had too perfect a childhood, does *that* not sound implausible? Moreover, if we do think this imagined individual with the perfect parents would be morally responsible, the

²⁰⁴ Kane, 1998, p 67

²⁰⁵ Except for a few special cases. Perhaps some people deliberate about how much money to spend on their mother’s birthday gift, and thinks that since their parents gave them *such a perfect childhood* she deserves the best there is. Still, the fact that their decision to buy a certain gift could have been *caused* by their upbringing remains irrelevant.

²⁰⁶ I will deal more thoroughly with the question of how one’s background may affect praise- and blameworthiness in Section 5.4.2.

same thing should be true about people living in a whole society consisting of such individuals.

Therefore, I cannot see my theory as forcing me to bite a bullet on Walden Two, since there is really no bullet to bite.

5.3.5. A Hardline Reply for Professor Plum

We have seen that it is not counter-intuitive to accept that people who have been manipulated to be kind and happy are morally responsible for their actions. However, what about people who are manipulated to do evil? In Pereboom's "four-stage manipulation case" there is a gradual move from a person who we allegedly all agree is devoid of moral responsibility, to a normal person in a deterministic universe. The idea is that we should see that every step along the way is morally irrelevant, so if the agent was not responsible in the first scenario, then he cannot be that in the last scenario either.

The story is about Professor Plum, who murders Ms White for selfish reasons. Plum can deliberate rationally about what to do, respond to reasons, is selfish but not compulsively so, and is overall described so that he will satisfy the most robust compatibilist freedom requirements. In the first scenario, Plum is directly manipulated. There are two versions of it; the original in which he is controlled from moment to moment by a evil scientists,²⁰⁷ and a new one in which the scientists merely push a button that causes Plum to reason in an egoistic way (this is not out of character for Plum, who often reasons in an egoistic way).²⁰⁸ In the second scenario, Plum was programmed at an earlier time to reason and deliberate the way he does, and then he murders Ms White.²⁰⁹ In the third scenario, there is no scientist but merely par-

²⁰⁷ Pereboom, 2001, pp 112-113

²⁰⁸ Kane, ed, 2011, p 410

²⁰⁹ Pereboom, 2001, pp 113-114

ents who brought Plum up to be the selfish murdering person he is,²¹⁰ and in the fourth scenario Plum is just a regular deterministic agent who commits murder.²¹¹ Pereboom's point is that Plum is obviously not responsible for murdering Ms White in the first scenario. Then for every new scenario something is changed, but according to Pereboom these changes are morally irrelevant features. It cannot matter if the evil scientist controls Plum from a temporal distance as well as a spatial distance, nor can it matter if it is a scientist or Plum's parents that control him, nor can it matter if the controller is a *person* at all or merely the forces of nature. Therefore, Plum is not responsible in any of the scenarios, according to Pereboom.

Michael McKenna has argued that one might as well go the other way around. Start with a scenario where Plum is just a normal agent who is clearly responsible. Then if every difference between the scenarios is morally irrelevant, one can draw the conclusion that Plum is responsible in the first scenario as well.²¹² McKenna also claims that the first case (in its original version) is under-described. Regardless of how the details are worked out, it is doubtful whether the case really supports the idea that Plum is not responsible. The idea of constant moment-to-moment control poses a dilemma: If the scientist always does whatever he pleases with Plum, Plum is not a person at all, but a mere dummy that the scientist plays with. However, if that is the case, it is not true that Plum (as has been stipulated) satisfies robust requirements for compatibilist free will – he can hardly do that if he is not even a person, but a mere dummy. On the other hand, if Plum is a person in his own right, with his own character traits, principles and preferences, the idea that the moment-to-moment manipulation undermines Plum's moral responsibility becomes less plausible. If Plum is to satisfy robust compatibilist requirements for free will, the scientists cannot just suddenly make him do something completely out of character. Often, they will just make Plum

²¹⁰ Ibid p 114

²¹¹ Ibid p 115

do what he would have done anyway, given his preferences and personality traits. Although they are still able to steer him somewhat, the intuition that Plum is obviously not responsible in this scenario loses much of its force.²¹³

Pereboom's new version of the scenario was developed in response to this criticism. His new and improved version is as follows: The scientist has the ability to momentarily cause Plum to reason in an egoistic way by pushing a button. Plum is often egoistic even without any manipulation from the scientists, so the button does not cause Plum to act out of character.²¹⁴ Here, it is undoubtedly the case that Plum is his own agent, not just a dummy, and that the scientist causes him to do something that he would not, or at least might not, have done anyway. Since Plum is often but not *always* egoistic when acting on his own accord, it might very well have been the case that Plum had not killed Ms White if the scientist had not given him that extra push in an egoistic direction. Now, as Pereboom himself points out, there are everyday situations where we momentarily are caused to become more egoistic by involuntary external influences. For instance, if the home team loses we become more irritable and egoistic.²¹⁵ Pereboom mentions this to explain how somebody can satisfy normal requirements for agency and be more than a dummy even in situations like these. However, reflecting on the similarity between the scientists pushing the egoist button on Plum and Plum's home team losing gives me the intuition that Plum's responsibility remains intact despite the button-pushing.

In the second stage of Pereboom's manipulation case, Plum is said to be *programmed* at an earlier time to act as he does. I think the word "programming" does some of the intuitive work here, since it leads us (despite the stipulated compatibilist free will of Plum) to imagine someone who is rigid,

²¹² McKenna, 2008, p 152

²¹³ Ibid pp 149-150

²¹⁴ Kane, ed, 2011, pp 410-411

²¹⁵ ibid p 411

unquestioning and unreflective. Pereboom imagines that the programming was made by scientists, in which case they were extraordinary lucky. Since they “programmed” Plum to be reflective, intelligent and so on, rather than to be a compulsive murderer, it must have been sheer luck on their part that Plum managed to encounter just those circumstances that eventually made him decide to murder Ms White. An alternative would be to have an omniscient (at least regarding the life of Plum) deity giving Plum just the right preferences and personality traits that, together with experiences the deity knew he would encounter, would make him murder Ms White.²¹⁶

According to PDR, the fact that a scientist tinkered with Plum’s brain at birth or a deity created Plum according to a certain plan is morally irrelevant. Neither fact could figure as a reason pro or con murdering Ms White, and neither fact makes deliberation and advice possible or impossible for Plum. Plum is stipulated to deliberate about his actions, and since he is not compulsive an adviser could still have something useful to tell him. Since Plum chose his actions under practical freedom, he is morally responsible for what he did. Is this counter-intuitive? I do not think so. Suppose one were to try Plum for the murder of Ms White, and learns that since Plum is reflective, can think rationally, question and argue about his own values and goals, and so on, there are absolutely no reasons whatsoever to excuse him on insanity grounds. However, then one discovers that when Plum was a baby, an evil scientist who wanted to see Ms White dead tinkered with Plum’s brain, and installed in him a selfish *streak*, no compulsive egoism by any means, but a selfish streak. The evil scientist hoped that Plum’s life would develop in a way such that he would end up killing Ms White because he could gain something from doing that, and the scientist was lucky. Would the installation of this selfish streak really strike anyone as an excuse that should get Plum off the hook? That seems extremely unlikely. I think most people would regard a selfish streak as much too weak to excuse him. If he had

²¹⁶ McKenna, 2008, pp 152-153 also suggests this as an additional step in the argument

been conditioned into a compulsive egoist it would be more plausible to excuse Plum and lay all blame on the scientist, but then Plum would no longer satisfy all compatibilist freedom requirements, and Pereboom's example would not work. Now suppose instead that we discover (somehow) that an omniscient deity had a special plan for Plum. He created Plum to be selfish but not compulsively so, to be rational, reflective, and so on, knowing full well that Plum would eventually encounter Ms White and murder her for his own selfish reasons. I do not think the fact that this deity is omniscient and had a plan for Plum alters things. The deity may know beforehand that Plum will encounter the right circumstances for murder while the scientist could only hope so, but it is still the case that Plum merely has a selfish streak, no compulsive selfishness. That should still be considered too weak to excuse him.

We now come to scenario three, where it was Plum's parents that caused him to be selfish. Since I have argued that it is not counter-intuitive to claim that Plum is responsible in the first and second scenario, I probably need not develop my thoughts on the third scenario in any length and detail. Suffice to say that I think it is easy to forget some of the stipulated facts about Plum – that he is rational, reflective, and so on – since on hearing that his upbringing determined him to be this way or that it is natural to associate to cult up-bringsings and other childhoods that cause a person to be *non*-reflective and *irrational*.

And if stages one to three are unproblematic, then obviously stage four, ordinary determinism, is as well.

5.3.6. A Hardline Reply for Alice

van Inwagen presents the following thought experiment to prove that *indeterminism* threatens the kind of free will required for moral responsibility as much as determinism does: Suppose that Alice is in a situation where she

deliberates about whether to lie or tell the truth. We suppose the world to be indeterministic; given the exact psychology Alice has and the exact circumstances she finds herself in, it is possible that she either lies or tells the truth. Alice eventually decides to tell the truth, and does so. Now, God rewinds the universe five minutes, and then, so to speak, presses the play button once again. We, the audience, who somehow witness this chain of events from outside the universe God is playing with, see Alice lying this time. God rewinds again, and again, and again, and sometimes Alice lies and sometimes she tells the truth.²¹⁷ In the audience some people have started to place bets on what Alice will do next. Now regardless of how Alice's brain works, if it follows some event-libertarian theory or if she has the power of agent causation, van Inwagen argues that we will be convinced that it is just a matter of luck whether she lies or tells the truth.²¹⁸ Since nothing explains why Alice, for instance, lied the third and the fourth time God pushed the play button, but told the truth the fifth time, it must be a matter of brute chance – and nobody can be responsible for things that happen by brute chance.²¹⁹

I think one problem with the claim that Alice lied or told the truth by chance, is that it relies on an intuition about a case that does not resemble anything we have ever encountered in real life. None of us has ever sat by God's side watching him rewind and replay the universe over and over again like a video-tape or DVD.²²⁰ What we *have* experienced on the other hand, are capricious people who change their mind all the time. One worry is that when we *think* we imagine the scenario as described, what we actually imag-

²¹⁷ Kane, ed, 2002, p 171. van Inwagen has repeated this argument in Campbell, ed, 2004, p 227-228, but there it is Marie who deliberates about whether to vote yes by raising her hand or not.

²¹⁸ *ibid* pp 172-174

²¹⁹ For another thought example with basically the same structure, see Alfred Mele, 2006, p 9. He uses different indeterministic worlds, perfectly similar up to a certain point in time, instead of rewinding and replaying one indeterministic world over and over again. It makes no difference to my argument whether we imagine different worlds rather than the same one being rewound.

²²⁰ See also Wilkes, 1988, for reasons to doubt our intuitions about unrealistic thought experiments.

ine is a capricious girl who changes her mind all the time, and this has a crucial effect on our intuitions. If a person is extremely capricious one might worry that there is something amiss with her rational thinking or that she has some psychological problem that undermines responsibility. Even so, let us suppose for the sake of argument that we are capable of accurately imagining even extremely far-fetched cases like this one, and then arrive at the intuition that Alice lies or tell the truth on a given occasion by chance.

I will grant that nobody can be responsible for things that happen by mere chance. At first sight there may seem to be counter examples. Suppose, for instance, that I happen to carry a rifle when I happen to see a person I really hate. I take aim and fire at him with the intention to kill. I am a bad shot, but I happen to hit him in the head and he dies. Suppose that everytime I have used the word “happen”, it signifies real chance. There will then be a lot of brute chance in this story, and yet I think most of us would agree that I am morally responsible for the murder. However, we do so because we presume that after I had seen the person I hate, and after the very idea that I might kill him had popped into my head, the fact that I decided to and tried to do this was non-random.²²¹ The problem with Alice is that she seems to make completely random *decisions*.

From a theoretical perspective (which is the perspective we have when we sit outside the universe by God’s side, watching Alice like a specimen in a Petri dish), Alice’s decisions look random, since we cannot give a causal explanation as to why she decides to lie in one replay and tell the truth in another. Let us then switch to a practical perspective, and discuss explanation in terms of reasons (qua reasons, not causes) instead. I have already explained that what causes one’s behaviour is irrelevant from the practical perspective. The same would be true of chance. If Alice finds herself in a

²²¹ See also Kane, 2002, p 418, for more examples where indeterminism does not preclude responsibility since the main character succeeded in doing what he attempted to do.

situation where she thinks it would be good to tell the truth, but she would also gain something from lying, she must decide which reasons are the weightier ones (from now on, I use “reason” in an internal, motivational sense – I do not suggest that there are external reasons that change from re-play to re-play). If Alice, before lying, thinks “Okay, lying would be a rather base thing to do here, but I do have so much to gain, and besides, everyone would do the same thing in my situation, and no one can ask of me that I should be a saint!” we have an explanation of her lying right there, in terms of reasons. I wrote earlier that I grant that no one can be responsible for what happens by mere chance, but Alice’s lying is not a matter of *mere* chance (although chance is all we see when we regard it from a theoretical perspective), since it is explicable by reasons. Her gain gives her a reason to lie, that truth-telling would be good gives her a reason to tell the truth, but the latter reason is weakened or outweighed by the fact that she has no reason to be a saint, and eventually the sum total of reasons point towards lying. In some other re-play of the universe, Alice may instead finish her train of thought by thinking “But telling the truth *is* the right thing to do after all!” and in that re-play, the fact that telling the truth was the right thing to do outweighed her other reasons, and once again we have a perfectly good reason-explanation of why she did what she did. As Carl Ginet writes in *On Action*, a reason-explanation is simply a statement that answers the question “why did the agent do that?”²²². That there exists a reason explanation of an action is compatible with there also being a causal explanation, even a deterministic causal explanation. But the former does not *require* the latter.²²³ I would say that reason explanations and causal ones are independent of each other, precisely because the first belongs to a practical and the second to a theoretical perspective.

²²² Ginet, 1990, p 130

²²³ *ibid* p 136

Donald Davidson has argued that a reason-explanation can only be an explanation at all if it has causal effect on the agent's behaviour, since there are clear cases where an agent had a reason to do something and also did this, but the action was caused by something other than the reason, and the reason therefore lacks explanatory value.²²⁴ I have already argued that an agent can have an efficacious will and practical freedom even if her mental life is an epiphenomenon. If she A:s if she decides to A and not-A:s if she decides to not-A, it is rational for her to deliberate about whether to A, regardless of epiphenomenalism. Suppose Anne lives in a universe where the mental is epiphenomenal. Her will is efficacious, but it is not the case that either thoughts or the exact physical correlates of thoughts cause behaviour. Anne now deliberates about whether to buy a DVD player or a CD player as a Christmas gift for her mother. She thinks that since her mother's old DVD player is slowly breaking down, and her mother loves to watch movies, she ought all things considered to buy a DVD player for her, and she does. Since I have already argued that it was rational of Anne to deliberate about what to buy, despite her mental life being epiphenomenal, it makes sense to also consider "mother's old DVD is breaking down, and she loves watching movies" a valid reason explanation of Anne's action, despite her thoughts about reasons having no causal effect on her actions. Now if we assume that her will was not efficacious either, she merely thought that it was, I am unsure whether we could still speak of a reason explanation for her action. Suppose that Anne would have bought the DVD player even if it were the case that it was her mother's CD player that was breaking down, and she loved listening to music. We might suppose that Anne has some kind of movie neurosis (perhaps she had a tough childhood in general, and the few fond memories she has are all movie-related) that would prompt her to buy the DVD player regardless of which reasons she thought of. I am inclined to believe that in a scenario where Anne does think of a good reason for buying the DVD player

²²⁴ Davidson, 2001, p 9

and then buys it, not knowing herself that she suffers from a neurosis that would have compelled her to buy it regardless of her reasons (that is, she *thinks* her will is efficacious), the reason still constitutes an explanation of her action, but I will not press this point. The important thing is that as long as her will *is* efficacious, as long as she will A if she decides to A and not-A if she decides to not-A, the reasons she thinks of before A-ing or not-A-ing constitute one kind of valid explanation of the action. There can thus be reason-explanations of an action even if the reason (or the agent's thought about the reason, or the physical correlate of that thought) did not cause the action. Going back to Alice and her lying/truth-telling, we must assume her will to be efficacious. If her lying and truth-telling was something that just happened to her, unrelated to her decisions, there would be an uncontroversial reason for her not being responsible for what she does, that does not hinge on the world being indeterministic (it could just as well have been a *seeming* randomness which in reality was a complicated kind of determination, and the judgement that she was not responsible would stand). Therefore, Alice's will is efficacious. When she decides to tell the truth, she does, and when she decides to lie, she does. Her lies can be explained by the fact that she could gain something from lying, and her truth-telling by the fact that telling the truth is the right thing to do, even though thinking through the same number of reasons pro and con the alternatives sometimes results in a decision to lie and sometimes to tell the truth.

Considered as a counter argument against PPC, van Inwagen's video-with-God argument once again supposes what ought to be proved. It supposes that the morally relevant perspective from which to judge a person free or unfree is the watching-video-with-God perspective. That any peculiar theoretical perspective of this kind would be the morally relevant perspective, is exactly what I question in this dissertation.

van Inwagen actually has another version of the argument from chance, where he views the problem from a practical perspective.²²⁵ In this version, he imagines himself as the main character of the thought experiment and questions whether he would really be in control of his actions. He still arrives at an incompatibilist conclusion, even though I have argued in this essay that one must take up a theoretical perspective to do so. I will now explain where I think van Inwagen's reasoning goes wrong.

van Inwagen imagines living in an indeterministic universe, where agents have the power to agent-cause things. He also knows (for some reason, perhaps God told him) that it is objectively possible that he either tells the press a damaging fact about his friend or that he keeps quiet. Since both scenarios are objectively possible, and the postulated objective probability that van Inwagen keeps silent is only 57 percent, he does not think that he is in a position to promise sincerely to keep silent.²²⁶ Thus, there is an element of chance here, but where does the chance come in? Remember my shooting example, where chance played an important part in the murder I committed, although it was not the kind of chance that seems responsibility-undermining. If van Inwagen's percentage came from there being a 43 percent chance that a journalist approached him and asked him about his friend, and van Inwagen was motivated to tell as soon as someone asked him (but could not be bothered to contact the press himself), he seems to be responsible for what he does. We could then say that he ought *firstly* to resolve to keep quiet and *secondly* to promise his friend to do so. This cannot be what van Inwagen means. The reason indeterministic Alice was *prima facie* not responsible for her lying or truth-telling was that the indeterminism occurred in deliberation, right before her decision-making. We must suppose the same thing here.

²²⁵ van Inwagen, 2000, pp 17-18

²²⁶ van Inwagen, 1983, pp 17-18

So there is a 57 percent chance that van Inwagen finds the reasons to keep quiet to outweigh the reasons to tell the press, and a 43 percent chance that he makes the opposite judgement. It now seems as if van Inwagen does not care that much about his friend's reputation, since it could quite easily be outweighed by other factors in his deliberation. Perhaps he is in no position to promise to keep quiet, simply because the promise would not be a sincere one. He ought *firstly* to care more about his friend and *secondly* promise to keep quiet. However, this does not seem to be the point van Inwagen is trying to make. van Inwagen wants to show that indeterminism itself would undermine promise-making, not that one ought to abstain from making insincere promises regarding things one does not really care about. As Alfred Mele has noted, if van Inwagen's argument is sound it should work in any indeterministic setting, even if the probability of keeping quiet is very close to one percent.²²⁷ Let us thus suppose that God tells van Inwagen that the objective probability of him keeping quiet is 99 percent. van Inwagen does care a lot about his friend, and in 99 percent of all possible worlds he will find his friend's reputation to outweigh all other considerations. In 1 percent of the worlds, however, van Inwagen will make an uncharacteristic decision and tell the press for a little bit of fame and attention. Now van Inwagen gets a call from his friend, who asks him to promise not to tell the press. What should van Inwagen do, knowing that in 1 percent of all possible worlds he will tell? Even if he decides not to promise anything, he is still left with the problem of deciding whether to tell the press. This decision cannot take the form "I will keep quiet in 99 percent of all possible worlds and talk in 1 percent" – that is simply no decision. In the same way as one cannot delegate decisions to the past and the laws of nature, or decide to do whatever one is determined to do, one cannot delegate decisions to chance, and decide to go

²²⁷ Mele, 2006, p 24

wherever indeterministic randomness leads one.²²⁸ Now suppose van Inwagen seriously decides to keep quiet. Why not also communicate that decision to his friend?

5.4. Arguments for Causal Compatibilism

PPC is opposed not only to incompatibilist theories, but also to those compatibilist theories that deny PDR and regard *some* causes of decisions and actions as responsibility-undermining. In Section 3.2.3 I discussed having more or less practical freedom, and at the end of 4.2.2. more or less causal freedom. When taking up a causal-explanatory perspective on people one could say that their actions were free in some minimal sense if the immediate causes are something in the agent's mind. They are free in a more substantive sense if the immediate mental causes of action are of the right kind (and what "the right kind" is may be different in different theories), even more free if they were agent-caused, and so on. I am not claiming that all free will theories that belong to this causal family can be neatly ordered along a scale. However, some compatibilist theories roughly belong somewhere in the middle; they demand something beyond that the agent's actions are caused by his own mentality, or his own desires, but they demand less than libertarian theories. Both Mele's and Fischer's compatibilist theories are of this type. While Fischer mainly argues against incompatibilists, Mele has argued extensively against so-called time-slice compatibilist theories, that is, theories according to which an agent can be morally responsible regardless of his or her history. PPC is a time-slice theory, since it says that all causes of behaviour are irrelevant, and thus a target for Mele's criticism. Mele supports his

²²⁸ It might seem like there is an obvious counter-example to this claim: I could decide to do whatever a die tells me, for instance. But if I threw I die to decide what to do, I would actually have made two active decisions: First I decided to throw the die, then I decided to let my action be regulated by the result. If I knew that there was some sub-atomic die-throwing going on in my brain, that there was some quantum randomness in my decision-making, I would still have to make active decisions, I could not just leave the decision to the quantum effects.

thesis by use of various thought experiments that I will discuss in the following section.

5.4.1. A Hardline Reply for Mele's Manipulation Cases

Mele thinks that one's values and traits must have been shaped in the right way for moral responsibility to be possible.²²⁹ ²³⁰ Mele agrees with compatibilists in general that there is an important difference between compulsion and causation. Caused acts one can be responsible for, compelled acts not. However, incompatibilists are fond of bringing up examples of people acting non-compelled, but from character traits and values that have been installed in them by brain-washers, to show that causation too is a threat. Mele responds that these are actually cases of compulsion, of being compelled to have certain values, certain traits, and so on. Actions that are caused by compelled values are not actions one can be responsible for either. A value is considered compelled if it fulfils the two conditions of being installed in a way that bypasses the agent's normal reflective and critical capacities, and if it was psychologically impossible to shed after it had been installed. Mele supposes that most regular people are not compelled to have the values they have in this way. Certainly, we acquire a lot of values as very young, when we have yet to acquire any "reflective and critical capacities", but these values are presumably "sheddable". That they are sheddable does not mean that the agent can get rid of the values in question given exactly how the past and the laws of nature are, since that would make Mele's theory incompatibilist. It just means that getting rid of them would not require any extraordinary circumstances. Only "unsheddable" values installed in us without our consent can deprive us of moral responsibility.²³¹

²²⁹ Mele, 2001, see for instance pp 170-171

²³⁰ Actually, Mele's primary interest is in autonomy, but he writes that a necessary although not sufficient condition for an agent being morally responsible for an act is that the agent was autonomous at an appropriate point in time. *ibid* p 140

²³¹ Mele, 2001, pp 153-156 and 167

To support his theory he uses thought experiments. All of these thought experiments involve people who have their values and/or personality traits completely transformed through an act of brainwashing that they had not consented to. Actually, Mele's so-called brainwashing does not resemble anything we can find in real life, but seems like an act of pure magic. In real life, brainwashing means to break people down by torture and/or isolation, until they are ready to accept whatever they are told (the results, however, are often not permanent – brainwashed people tend to go back to their old views once restored to a safe environment).²³² In fiction, brainwashing tends to be associated with criminal master-minds and cult leaders. However, in Mele's thought experiments, it is stipulated that the so-called brainwashing consists of changing some values or goals a person have, without him or her noticing. He or she will be just as happy, intelligent, harmonious and rational after the brainwashing as before it. I think it might be the case that the word "brainwashing" is doing some of the intuitive work here, since it carries such negative connotations. Therefore, I will use the word "magic" instead.

Now, let us examine Mele's example. The first one is about Beth, who used to value things other than philosophy, but is magically turned into an extremely dedicated philosopher who wants nothing more than philosophising all day long. The second example also features Beth, but here she is turned into an evil murderer (although still rational, critical, intelligent, and so on). The third example features Charles Manson who is turned into a kind and benevolent person. By hypothesis, the examples of Beth 1, Beth 2 and Manson fulfil everything a compatibilist could ask for after the respective conversions, and this, of course, includes having practical freedom. According to PPC they are moral agents and have moral responsibility for what they do. Mele suggests that this is counterintuitive, but I disagree.

Let us first look at good and benevolent Manson. Suppose he does some good deed, such as saving a child from being run over by a car. Does he not

²³² Myers, 2002, pp 145-146

deserve to be praised for this action? Suppose Manson and his neighbour (with a normal background) both rescue a child from being overrun by a car. Should we perhaps praise them both, but for different reasons? The neighbour because he deserves it, and Manson because we find him useful and want to encourage him to behave well in the future? Should we, in short, think of Manson's neighbour as a real human being and the just as intelligent, rational and benevolent Manson as a useful device? When keeping firmly in mind that Manson is as rational and coherent as anybody; has a stable, benevolent and virtuous character; and rescues the child because he truly cares more about other people than about his own safety, I find no plausibility at all in denying him the same amount of and the same kind of praise as anybody else who does a good deed. To treat an intelligent and benevolent human being as a mere tool or trained animal, to praise him simply in order to reinforce good behaviour while praising his neighbour in earnest, seems to me the counterintuitive alternative. At least free will sceptics like Pereboom, who wants to alter our praxis of praise and blame, claim that it should be altered for everybody. However, to praise Manson's neighbour because he deserves it, while simultaneously praising Manson as a mere training device, seems downright disrespectful and unfair towards the latter.

When it comes to Beth 1, the industrious philosopher, it is also hard to see any justification for disregarding her as a moral agent or as responsible for what she does. Suppose she produces a real master-piece in her field. Does she not deserve to receive royalties for the book? I wrote earlier that the question of whether an agent is morally responsible should not be confused with the question of how we ought to treat people, and somebody could believe that there are various non-responsibility-related reasons for giving or not giving royalties to people. Suppose, however, for the sake of argument, that we think people in general *deserve* to receive royalties for the books they produce, and therefore ought to be given royalties. Should this

judgement not extend to Beth 1? Should the book be regarded as not produced by anybody in particular, as similar to something great that natural forces have accomplished? Perhaps we should think of Beth 1 as we think of a rescue dog? A rescue dog cannot really be morally responsible for saving people's lives, but we still reward these dogs with treats and toys in order to reinforce the rescuing behaviour. Likewise we might reward Beth 1 with royalties simply in order to reinforce her philosophy writing behaviour, while we give royalties to other philosophers because we think they deserve it. Just as with Manson and his neighbour, I find it deeply counterintuitive that Beth 1 should be either treated differently than her colleagues or treated the same when it comes to outward behaviour but with a completely different justification.

When it comes to Beth 2, who was turned into a murderer, it is of course a great pity that this happened – perhaps more of a pity than if somebody grow up to be evil without ever living a good life. In the case of Beth 2, one may not only consider her current evil unfortunate, but also miss her former, benevolent self. But do these considerations have anything to do with the question of moral responsibility? I do not think so. If Beth 2 and another, equally evil but simultaneously intelligent and rational woman, committed some heinous deeds together, I actually find it very counterintuitive to treat them differently. Keep in mind that they are both *completely rational and coherent*. Keep in mind that they *did this together*; it was in no way Beth 2's accomplice who coaxed or dragged her into the situation. Under these circumstances, would it not be terribly odd to place Beth 2 in a facility for the criminally insane and her accomplice in jail? Would it not be terribly odd to hate her accomplice but simply pity the just as intelligent, coherent and evil Beth 2? If one hesitates to put Beth 2 through retributive punishment, I think one should carefully investigate one's intuitions and consider whether this is not because one has misgivings about retributive punishment in general. As I pointed out earlier, one does not have to endorse retributive punishments just

because one believes in desert-entailing moral responsibility. I think there is a real risk that one confuses non-retributive intuitions with intuitions about moral responsibility if one focuses only on manipulated individuals. Therefore, let us focus as much on Beth 2's accomplice; we may call her Carol. Carol and Beth are very much alike. They are intelligent, rational, and profoundly uncaring. The only difference between them is that Beth 2's personality was created by a wizard, while Carol has a normal background. Nowadays, however, one cannot tell them apart. They commit some heinous crimes and are caught by the police. Now Beth 2, who was just as active and delighted in the criminal life as much as Carol did, is whisked away to a facility for the criminally insane; this, Carol is told, is standard procedure for all people whose personality was once created by a wizard. It does not matter that Beth 2 is just as sane as Carol. Carol, however, will be put in the most horrible jail in the country. No good consequences will come out of this; no one will be deterred. It is simply considered *fair* to split up the criminal psychological twins and put one of them in an asylum to be cared for and another in prison to suffer. For my own part, I find all this very counter-intuitive. If we think that Beth should not be made to suffer, simply locked up to protect the rest of society, we should extend that judgement to her psychological twin Carol.

Somebody might object that if Manson is morally responsible for his actions after being turned good, this would mean that we should not just reward him for saving the child, but also punish him for crimes he committed before the manipulation, and this is counter-intuitive. However, an agent can only deserve punishment for crimes he committed himself, and I will argue that Manson is a *different moral agent* after the conversion.

A moral agent, according to PPC, is somebody who has practical freedom, which means he or she must choose his or her actions and believes he or she has an efficacious will. This is the definition we need to keep in mind

when asking ourselves whether Good Manson at T2 is the same moral agent as Bad Manson at T1. Did Bad Manson choose Good Manson's actions, and did Bad Manson believe that his will was efficacious in this regard? The latter part of the question is difficult to answer, since Bad Manson might have believed that his will was efficacious when it came to choosing the future actions of Manson simply because he did not realise he would later undergo such a profound change. It is probably the case that it should be answered in the negative; if Bad Manson did not know that Good Manson would be created, he would not believe that he could choose for Good Manson either, and if he had known and understood the magnitude of the change he would later undergo, he might very well have believed his will inefficacious. In any case, the first part of the conjunction is clearly false, at least when it comes to morally important choices. Bad Manson did not choose what Good Manson would do. Bad Manson might have planned all kinds of dark deeds, but as soon as the wizard waved his magic wand, these plans fell apart. There may be continuity between Good Manson and Bad Manson regarding memories, and perhaps regarding small and unimportant choices as well (perhaps Bad Manson had decided to buy a loaf of bread the next day, and after the magic conversion, he still carried out that decision since he still needed bread), but there is no continuity at all when it comes to morally important choices. Therefore, Bad Manson and Good Manson should be considered *different moral agents*. Since a moral agent can only be morally responsible for what the same moral agent has done, Good Manson cannot be morally responsible for what Bad Manson did.²³³

How does moral agency relate to personhood then? If Bad Manson and Good Manson are different moral agents, does this mean that they are also different persons? This question seems to depend on what theory of person-

²³³ Those readers interested in the subject of personal identity could read more in Korsgaard, 2008. In chapter 13, "Personal Identity and the Unity of Agency: A Kantian Response", pp 363-397, Korsgaard addresses the issue of personal identity from a practical perspective, in a way that I take to be compatible with what little I have written here.

hood we subscribe to. If a person is an animal, a living organism, it would follow that Bad Manson and Good Manson are different moral agents despite being the same person. This seems problematic. On the other hand, if personhood is tied to mental characteristics, it is possible that they are different persons as well as different moral agents. There is continuity in memories, that is true, but regarding their values and personality traits there is a sharp, clean break at the time of manipulation. However, to delve deeply into the question of what makes A and B the same or different persons lies outside the scope of this dissertation. I have provided a definition of what it takes to be the same moral agent. I suggest that personhood should accompany moral agency, but I will not go into details.

Once again, I have shown that my theory does not force me to accept counter-intuitive implications. Objecting to Mele's view is not counterintuitive at all. But even if Mele, or some other causal compatibilist with a "historical" as opposed to "time-slice" theory, would come up with a thought example that forced me to accept a counter-intuitive conclusion, I would do so, since I think PPC and its prime argument PDR are more trustworthy than moral intuitions on far-fetched thought experiments.

5.4.2. Kaye's Mitigating Factors

Mele's arguments are about purely hypothetical cases. He may have intended them to cover some real-life cases as well – an example of a father teaching his child various religious beliefs indicate as much. There is no evidence, however, that it would be possible in real life to compel somebody to acquire certain values while leaving him completely intelligent, rational, and reflective. Anders Kaye, on the other hand, has argued that not just magical brainwashing but also mundane real-life phenomena like bad childhoods could undermine responsibility, even if they do not compromise a person's capacity for critical thinking. Kaye is an incompatibilist (although he consistently use the word "originationist" instead), but the arguments I will ex-

amine here attempt to support the importance of an agent's history and are therefore better treated under the "causal compatibilism" heading. I will briefly say something about why his arguments fail to support incompatibilism or libertarianism before discussing why they do not even prove that one's history is important in itself.

I will grant Kaye that we (or most people, or most sensible people, or most people who have reflected properly over the matter) think that a bad childhood is relevant when judging criminal behaviour. Kaye thinks this is best explained by the idea that one cannot be responsible for actions that were caused by factors beyond one's control.²³⁴ He argues that this is the case with criminals from bad childhoods; these people were caused,²³⁵ *determined* even,²³⁶ by their childhood experiences to become criminals. Kaye thinks that criminals formed by difficult social conditions does not have a genuinely free will, since their acts were formed by factors beyond their control.²³⁷ He further argues that, because we should excuse criminals with a bad childhood, the originationist approach is attractive.²³⁸

The claim that people in general have libertarian free will while criminals from harsh conditions are determined is very problematic. No matter how one defines difficult social conditions, it does not correlate perfectly with criminal behaviour. Difficult social conditions do not in themselves *determine* that somebody will become a criminal. Some very specific difficult social conditions, together with other environmental influences and genetics, might determine that somebody will become a criminal, but if so, we seem to approach a more general deterministic thesis, according to which all people

²³⁴ Kaye, 2007, pp 372 and 403-405

²³⁵ *ibid* p 403-404

²³⁶ *ibid* p 372. Kaye equates causal with determinist when he discusses the originationist account of excuses. Since he hold that hard social conditions are excuses on this account, it would follow that they determine people to become criminals.

²³⁷ *Ibid* p 369 on the originationist view on responsibility and free will, and p 397 on how difficult social conditions cause certain dispositions and feelings, and disdain for social norms and authority, and p 396 on how difficult social conditions catalyse antisocial conduct.

²³⁸ *Ibid* p 373

are determined to do what they do by genes and their environment. However, it would be terribly odd if some people were determined by the past and the laws of nature to act as they do while others were undetermined, a claim that according to Kaye most people believe, and one which he seems to find at least plausible himself.²³⁹ If most adults have a libertarian free will, it is not plausible that difficult social conditions can wipe that out except by means of destroying cognitive capacities. Simply explaining correlations between difficult social conditions and criminal behaviour is no problem for a libertarian – libertarian Tim O'Connor, for instance, thinks that people have *different dispositions* that makes certain actions much more *likely* than others for a given person.²⁴⁰ This suffices to explain correlations between background and criminality. If pleasant social conditions (together with some other environmental factors and genes) cause disposition A, which makes it *unlikely* that one will perform criminal acts, while difficult social conditions (together with some other environmental factors and genes) cause disposition B, which makes it *likely* that one will become a criminal, this explains correlations. If I have a strong criminal disposition, honesty may be tough for me, but tough is different from impossible. (In rare and extreme cases, the disposition could, of course, be so strong that the person is compelled to follow it.) The view that all normal adults have libertarian free will is thus consistent with there being a correlation between difficult social conditions and criminal behaviour.

Kaye thinks a theory need to be sensitive to the history of the agent, if difficult social conditions is to be taken into account.²⁴¹ But the time-slice theory of PPC is actually well equipped to explain how difficult social con-

²³⁹ Ibid p 373

²⁴⁰ O'Connor, 2002, pp 97-98

²⁴¹ *ibid* p 382. "Ascertaining whether an actor has these features and opportunities requires an inquiry into a narrow set of facts – facts that can be captured by a "time-slice" "snapshot" of the actor at the time of the act itself. Such snapshot feature/opportunity inquiries of course tell us some important things about the actor, but they provide little opportunity to incorporate hard social conditions into the responsibility determination."

ditions can be a mitigating factor, if only indirectly. Finding out correlations between difficult social conditions and criminal behaviour might have been what first prompted people to discuss mitigating factors, but they can only be a first step. A natural response for the hardline retributivist would be to point out that these are not 100 percent correlations. Pre-theoretically, many people are inclined to view Chris, a criminal from difficult social conditions, as less (or not at all) responsible for his actions than Richie, an honest man who has lived a thoroughly pleasant life. But if we bring Harry, an honest man with a background just as difficult as Chris', into the equation, this will probably affect our intuitions. If Chris and Harry have similar backgrounds, then why cannot Chris be as good as Harry? What is his defence?

I argued in 3.2. that practical freedom comes in degrees, and that we only have maximum practical freedom when our values are fully coherent, and we do what is best according to them. Now perhaps we will find that Chris suffers from something that straightforwardly diminishes his practical freedom. His difficult background (together with genetic disposition and other environmental influences) may have caused him to leave certain options unconsidered, options that are better according to his own values than the actions he actually performs. Having unconsidered options of this kind is freedom-undermining according to PPC. Suppose, for example, that Chris dislikes the dangers involved in living a criminal life, and envies the predictability at which monthly salaries arrive for people with steady jobs. Suppose he thinks that non-criminals with steady jobs have better lives than he has. Suppose also that he has the relevant information on how to go about getting a normal job and/or getting a proper education, and he believes his will to be efficacious. He might not consciously think "if I decided to apply to this school, I could", but he does not believe that anything would prevent him from doing so if he tried. There are no outer obstacles forcing Chris into a criminal life. According to PPC, his freedom could still be diminished, if he does not consider turning honest when *deliberating* (rather than merely en-

gaging in first-person fantasies) about what to do, despite this being better than criminality according to his own values. Perhaps his unfreedom goes deeper than this; perhaps some of his values were chosen under much diminished practical freedom. I think it is highly plausible that many people who grew up in bad circumstances do have a diminished practical freedom in this way. Since Harry, with a background just like Chris's, did become honest, Harry must have seen honesty as an option, but this does not mean that Chris necessarily did. Just because their backgrounds are the same, they do not have to be psychological twins. If this is the explanation of Harry's honesty and Chris's criminality, that Harry saw a real option where Chris could not see one, Harry has more freedom than Chris. Chris would thus have less responsibility for his criminality than Harry has for his honesty.²⁴²

Now let us instead, for the sake of argument, suppose that Chris and Harry are psychologically very similar, aside from having similar backgrounds.²⁴³ They both thought things through, and then Chris decided to become a criminal, while Harry decided to be honest. If they both had close to full practical freedom, they must both be morally responsible to a high degree. This also accords well with my own intuitions at least. In real life, we may have to rely on rules of thumb, since we never have full information about people's psychology, and one such rule of thumb might be to regard

²⁴² An asymmetric Wolfish view is actually not implied here, since according to my view the opposite is possible as well. If somebody held values according to which a life of crime would be better than honesty, but due to his completely honest environment he could not really consider becoming a criminal, then he would only have a mitigated amount of responsibility for his honesty. I suspect though that most honest people does not hold values according to which criminality would be superior, and if so, they are fully responsible for being honest. If they are not praised for their honesty, that is because we normally only praise people for doing *difficult* things, not for doing things that come naturally – but one may still have full responsibility for doing something easy, even if it does not merit any particular praise. It might also be the case, at least in theory, that somebody is a fully responsible evil criminal who chose this way of life under perfect practical freedom. However, as I have pointed out earlier, I think this rarely or never happens in real life.

²⁴³ With similar character traits and values to start with, they are both, for example, reflective and think things through properly before they act. I obviously do not mean to stipulate that they are so similar in every mental detail that it becomes trivial that they will always act the same.

people from a difficult background as having diminished responsibility for their actions. However, picture someone who, despite a difficult background, made a carefully considered decision to become a criminal, who would like nothing better in life. My own intuitions are that this person is just as responsible for his or her actions as a criminal from a pleasant background.

However, this does not necessarily imply that the amount of praise that Harry deserves is equal to the amount of blame that Chris deserves. It is a fairly common assumption, one which can easily be incorporated within PPC even if it does not follow from it, that one deserves more praise for doing the right thing when doing so is difficult, and conversely that one deserves more blame for *not* doing the right thing when doing so would have been easy.²⁴⁴ If one extremely good and brave swimmer and one bad and fearful swimmer both rescue a child from drowning in a deep lake, the latter swimmer deserves more praise. If a soldier refuses to torture a prisoner despite it being expected of him, he deserves praise, while I deserve no praise at all for not torturing people. If a soldier does torture somebody in a situation where this is what everybody expects, he is at least arguably less blameworthy than a regular person who tortures somebody without this social pressure.²⁴⁵

Now we may assume that it is the case for both Chris and Harry that going for an honest life would involve a lot of struggle, while becoming a criminal would have been rather easy. And it is also plausibly the case for people with pleasant social backgrounds that being mostly honest seems obvious to them and does not require any particular effort, while embarking on a criminal career would have been difficult and required some determination. If that is so, it means that Harry deserves praise for his honesty, while your average honest person may deserve none, and simultaneously Chris deserves less

²⁴⁴ Since how difficult something would be is typically relevant to deliberation it does not contradict PDR.

²⁴⁵ All these examples are of course supposed to include an “everything else being equal” clause.

blame for his criminality than a criminal person from a pleasant background would.

5.5. Summary of Chapter 5

In Chapter 3, I argued *for* PPC. In this chapter, I responded to various counter-arguments *against* PPC – or rather to groups of theories where PPC is included, compatibilism and time-slice compatibilism. Formal arguments for incompatibilism, as well as thought experiments designed to elicit incompatibilist or history-sensitive compatibilist intuitions, fail as counter-arguments against PPC for the same reason. They all assume what ought to be proved. I support PPC with PDR, the principle of deliberative relevance, according to which facts that are irrelevant for a deliberating agent or adviser cannot be morally relevant. I have also argued for PDR. However, the formal arguments and the thought experiments simply assume that one *should* focus on causes of the agent's behaviour that cannot figure as reasons in deliberation or advice. They assume that PDR is false, without arguing for that assumption.

Perhaps one should reconsider PDR if the acceptance of this principle would lead to extremely counter-intuitive results regarding some well-known thought experiments in the debate. (I write “perhaps”, since it is not obvious how reliable our intuitions are regarding far-fetched scenarios that do not resemble anything we have encountered in real life.) I have shown that this is not so. At least my own intuitions accord well with the conclusions that can be drawn from PDR and PPC regarding moral responsibility in various thought experiments. There will of course be readers whose intuitions differ from mine. Still, I think I have shown that PDR does not imply that we have to accept any strongly counter-intuitive conclusions.

Since I have given strong arguments *for* PDR, since PDR supports PPC, since PDR does not lead to any strongly counter-intuitive conclusions and

since all arguments against PPC begs the question by simply assuming that PDR is false, I conclude that we ought to accept PPC.

6. Moral Agency, Responsibility and Metaethics

Thus far I have argued for a specific form of compatibilism, PPC, by an appeal to PDR. By showing that incompatibilist arguments cannot undermine PPC, since they must assume the falsity of PDR, I have argued that these incompatibilist arguments need to be addressed regardless of whether one is a deontologist or utilitarian on the normative-ethical level. However, if we move up to the metaethical level, PPC gains extra plausibility by being compatible with any metaethical theory. It is not, on the other hand, the case that incompatibilism or causal compatibilism is compatible with any such theory.

6.1. What Determinism Can and Cannot Imply

It has often been pointed out that although some kinds of freedom may be incompatible with determinism, others are obviously compatible with it. Determinism may mean that no one ever could do otherwise in some ultimate sense, but it cannot mean that no one ever does what she wants, or that no one ever is what Frankfurt calls whole-hearted. Likewise, determinism cannot mean that no one ever has a pro-attitude to someone else's action, or that no one ever does what is endorsed by the society in which he or she lives. So if "P did what was morally right to do" means something like "P did something that I have a pro-attitude towards" or "P did something that is embraced by our society", then P can do right (and, correspondingly, do wrong) regardless of whether the world is deterministic or indeterministic.

This simple, metaethical point is often overlooked in the debate. However, a few philosophers have used metaethical arguments to support compatibilism.

6.2. Korsgaard and Constructivism

Korsgaard considers practical freedom sufficient for moral agency because she is a constructivist. Since we have practical freedom, we must choose what to do. Since we must choose what to do, we need reasons for choosing one alternative over another. Reasons for action stem from so-called practical identities – being a wife, a dog-owner, a PhD student each provides me with reasons for action in various circumstances.²⁴⁶ Many of these practical identities are sheddable, and only provide us with reasons for action given that we care about them. Being a PhD student gives me reason for writing on my dissertation, as long as I care about being a PhD student, but cease to be reason-giving if I decide that I would rather have another job. However, being a human being, a rational animal, is non-optional, and from our rational agency and animal nature stems reasons – moral reasons – that hold unconditionally.²⁴⁷ Korsgaard then argues that reasons for action are inherently shareable and public rather than private. This argument leads her to reject the idea that my humanity gives me reason to promote my interests while other people's humanity gives them reason to promote theirs, favouring instead an altruistic morality.²⁴⁸

This moral theory implies that practical freedom is sufficient freedom-wise for moral agency. Reasons (moral as well as others), on this view, are not mind-independent entities that we discover, but something we have because we must think in terms of reasons when deciding what to do. This is so because we have practical freedom. Whether we also have freedom of a libertarian kind, or whether our values have been caused in the right way,

²⁴⁶ Korsgaard, 1996a, p 101

²⁴⁷ *ibid*, pp 120-121 and 152-153

²⁴⁸ *ibid*, pp 132-145

makes no difference for our need of reasons. One can of course question Korsgaard's account of morality on several points. It is true that I cannot choose whether to be a human, a rational animal, or not. But does this fact really imply, as Korsgaard thinks, that I must offer respect, help, and so on to other rational agents and other animals? That depends on, among other things, exactly how one understands the "must" here, and whether her argument for reasons being essentially public rather than private is valid. What it does not depend on is whether the universe is deterministic, or the details of my personal history. If Korsgaard's arguments are valid, the existence of moral agency can be inferred from the fact that there are rational agents with practical freedom. If so, it follows that practical freedom is sufficient for moral agency, and moral agency is compatible with determinism.

Sharon Street, in her "What is Constructivism in Ethics and Metaethics", presents the following quote as the canonical characterization of constructivism:

[T]he constructivist is a hypothetical proceduralist. He endorses some hypothetical procedure as determining which principles constitute valid standards of morality ... [He] maintains that there are no moral facts independent of the finding that a certain hypothetical procedure would have such and such an upshot.²⁴⁹

Normative truth is not understood as being merely uncovered by or coinciding with the outcome of a certain procedure, but as constituted by emergence from that procedure.²⁵⁰ However, Street thinks this canonical characterisation fails to capture what is really constructivism's core. She prefers instead a standpoint characterisation. What is special about normative truths, according to Street, is that they can only be arrived at from a practical point of

²⁴⁹ Darwall, Gibbard and Railton, 1992, p 115

²⁵⁰ Street, 2010, pp 365

view.²⁵¹ She defines the practical point of view as that of somebody who *values* things.²⁵² “The subject matter of ethics is the subject matter of what follows from within the standpoint of creatures who are already taking this, that, or the other thing to be valuable”.²⁵³ What does it mean that something follows from within a standpoint? Plausibly, it means that it follows from some propositions I must accept when taking up this standpoint. If the practical standpoint is the standpoint of somebody who values things, I must, for instance, accept the proposition “some things are valuable” when taking up this standpoint. Anything entailed by that proposition would then be a normative truth.

Now, Street distinguishes between Kantian and Humean metaethical constructivists. Kantian constructivists believe that moral conclusions follow from within the practical standpoint when it is given a formal characterisation, while Humeans think one has to put substance in to get substance out.²⁵⁴ Korsgaard is a constructivist of the Kantian kind, since she believes that a substantial morality can be inferred from our practical freedom, the essentially public nature of reasons and our nature as rational animals. Humean constructivists on the other hand, believe that one cannot infer a substantial morality from anything less than substantial assumptions about *what it is* that is valuable. Suppose it is simply the case that from a practical standpoint I take happiness, respect and cooperation to be valuable. Given this, it might be possible to infer substantial moral principles from the propositions I must accept.²⁵⁵

²⁵¹ *ibid*, pp 365-366

²⁵² Street’s characterization of the practical viewpoint is thus slightly different from but compatible with my characterization of the practical perspective. I write that the practical perspective is the perspective of a deliberating agent or an adviser, and that in choice we agents try to pick the option which is best according to their own values. This means that the practical perspective does include valuing things.

²⁵³ *ibid*, p 367

²⁵⁴ *ibid*, p 369

²⁵⁵ *ibid* p 370

I wrote that Korsgaard's constructivism implies that practical freedom suffices for moral agency. This is the case for other Kantian constructivist theories as well. I have already argued that an agent with practical freedom tries to choose the option that is best according to his own values – which means that all agents with practical freedom are valuers. If moral principles can be inferred from propositions that all valuers must accept, it follows that moral principles can be inferred from propositions that all agents with practical freedom must accept. Agents with practical freedom would be bound by moral principles, and therefore be moral agents. If all moral truths consist of principles following from propositions that valuers must accept, this includes truths about moral responsibility. Whether agents with practical freedom are morally responsible as well will thus depend on whether principles incorporating the concept of moral responsibility follow from propositions that all valuers must accept. It does not, however, depend on whether the universe is deterministic. The truth of determinism could not affect an entailment relation. If A implies B, then A and determinism or A and indeterminism also imply B. If a proposition according to which people are morally responsible for what they do follows from propositions all valuers must accept, it does so regardless of determinism or indeterminism. Therefore, Kantian constructivism implies compatibilism on moral responsibility as well as moral agency.

According to a Humean constructivist theory it is also the case that moral agency is compatible with determinism. If actions are right and wrong depending on how they conform to moral principles that are entailed by propositions one must accept if one is a valuer who values (for instance) well-being, respect and cooperation, it follows that actions can be right or wrong regardless of determinism, since there can be valuers who value these things in a deterministic universe. It also follows that moral responsibility is *possible* under determinism; it could be the case that given the values of a society of agents, a system of moral principles incorporating the concept of moral responsibility follows. But what if everybody valued a decidedly incompati-

bilist idea of moral responsibility? We could imagine that all agents value moral responsibility and desert according to Galen Strawson's conception of these matters, where one has to be a *causa sui* to deserve praise or blame,²⁵⁶ while no other idea of moral responsibility and desert is either entailed by value propositions the agents accept or find valuable in themselves. If that were the case, and Humean constructivism were the true metaethical theory, it would follow that there is no moral responsibility or desert in the world. It is still not obvious that incompatibilism would be true in this scenario. The non-existence of moral responsibility and desert would depend not only on the fact that the universe must be either deterministic or indeterministic (and both rule out the possibility of a *causa sui*), but also on the values that all agents happen to hold. If we take "incompatibilism" to mean that determinism (or indeterminism) *in itself* rules out desert-entailing moral responsibility, incompatibilism would still be false in this scenario. On the other hand, if we take "incompatibilism" to mean that determinism (or indeterminism) rules out moral responsibility given the contingent facts about the world, we would have incompatibilism in this scenario. There is thus no straightforward implication from Humean constructivism to compatibilism. While Kantian constructivism implies compatibilism, Humean constructivism merely suggests it.

6.3. Contractualism

James Lenman has argued for compatibilism on metaethical grounds in his "Compatibilism and Contractualism: The Possibility of Moral Responsibility". He wants to argue, against Erin Kelly, that moral responsibility and desert can have a place in morality even if the world is deterministic,²⁵⁷ and,

²⁵⁶ I assume this for the sake of discussion, but it is unlikely to be the case. See Nahmias, Coates and Kvaran, 2007; Many people seem to consider desert possible under determinism.

²⁵⁷ Lenman, 2006, pp 8-9, pagination from the printed version from <http://www.jstor.org/stable/10.1086/508035>

against Scanlon, that moral responsibility is more than just attributability. The claim that an act was morally wrong plausibly involves, according to Lenman, that some sort of sanction is justified.²⁵⁸ Lenman argues that compatibilism follows if we accept a contractualist account of morality, according to which moral principles are those we could not reasonably reject. The metaethical ideas in Lenman's article are actually fairly close to what Street calls "Humean constructivism". We are agents who must decide what to do and social animals who must find a way to live together. There are certain things we simply find ourselves valuing, and these values determine which shape the answer may take. Now, what Lenman attempts to establish in his article is that *at least* idealised human beings with a high degree of both coherence among their values and beliefs and a high degree of self-control (in a compatibilist sense) could be morally responsible for their actions even if determinism were true – even under determinism it could be the case that it is not reasonable to reject moral principles according to which people are morally responsible for what they do.²⁵⁹ If he succeeds, he will have proven compatibilism. He then leaves it open as to how close real human beings come to his idealised ones, and to what extent real humans can be responsible.

When trying to decide whether or not it is legitimate to hold people responsible we should ask ourselves whether we could accept or reject a moral principle that says that we ought to do this. Lenman's idealised human beings accept this principle because they respect each other. Holding each other responsible is to accord each other a form of respect.²⁶⁰ Now an incompatibilist might think that respecting each other in this way requires seeing each other as ultimate originators of causal chains or the like, but Lenman argues that this is irrelevant. His idealised human beings are, as already mentioned, coherent and in (compatibilist) control of themselves. From the perspective

²⁵⁸ *ibid*, p 4

²⁵⁹ *ibid* p 11

of a deliberating agent, his or her future actions will thus be up to him or her, and they will be actions that he or she considers good and is willing to embrace.²⁶¹ It is therefore reasonable of him or her to accept a principle according to which he or she will be responsible for these actions. Granting respect to others by holding them responsible for their actions merely acknowledges that they also are coherent and in control of themselves, and thus they have the same reasons as he or she has for accepting a principle according to which they ought to hold each other responsible for what they do. Accepting a principle of holding each other responsible in this strong sense does carry a cost, since people who do wrong are liable to sanctions.²⁶² It is worth it still, however, for the above given reasons.

One might still worry that the moral responsibility Lenman is discussing is not of the desert-entailing kind. It seems like he intends it to be, since he wants to answer the worries of Kelly, and what she argues is that *desert* and the corresponding moral responsibility concept cannot be made sense of. That he really is discussing the desert-entailing concept becomes clear when he argues that the claim that an action was morally wrong *involves* that a sanction is justified. As I have mentioned before, desert is a three-part relation between agent, action and praise/blame. If a theory on moral responsibility holds that something outside this triad, like the consequences of holding responsible, is what justifies the praxis, then the theory is not about desert-entailing moral responsibility. Lenman, however, thinks that the moral wrongness of an action in itself justifies a sanction, which means that the desert-relation holds between agent, action and blame/sanction.²⁶³ It is precisely this kind of moral responsibility, not a weaker or less demanding con-

²⁶⁰ *ibid* p 12

²⁶¹ *ibid* p 16-17

²⁶² *ibid* p 12

²⁶³ This does not mean that Lenman thinks there can never be excuses. He holds, with Wallace, that an action is excused if the individual did not really do anything wrong. *Ibid*, p 7. Circumstances where we excuse are cases where the agent charged with some blameworthy

cept, that it would be reasonable for us to accept, and which therefore has a place in morality on Lenman's contractualist account. (Lenman adds that there could also be good utilitarian reasons for holding each other responsible, but he does not think they are sufficient in themselves for grounding moral responsibility.²⁶⁴)

Lenman's argument would not convince somebody like Pereboom, who believes that it would be possible to rid ourselves of all ideas of desert and desert-entailing moral responsibility, and that we would be better off without them. Nonetheless, as long as we think morality consists of whatever principles it is reasonable to accept in order to be able to live together, the truth of determinism (or indeterminism) cannot in itself rule out moral responsibility. If holding somebody responsible really is to show this person respect, if we value respect, and if we are coherent and controlled enough to perform actions we consider good, it seems that we do have reason to accept a responsibility-holding principle, regardless of whether determinism is true. Now, somebody might object that the interesting thing in the free will debate is not whether we have reason to accept certain principles, but whether there truly *is*, independent of our principles, something like desert or moral responsibility. However, that objection is simply an objection towards Lenman's metaethics – a rejection of contractualism and a demand for realism.

6.4. Peter Strawson's Naturalism

In his famous article "Freedom and Resentment", Peter Strawson advances several arguments in support of the praxis of holding each other responsible. One is that we could not rid ourselves of this praxis even if we tried to, it is simply too deeply entrenched in our psychology, and it is pointless to argue

action turns out not to have performed the action intentionally at all – at least not under the description under which it is blameworthy.

²⁶⁴ *ibid*, p 18

that we should do something that we cannot do.²⁶⁵ Another is that we should deem on reflection that our lives are made richer by this praxis, and therefore we should stick to it.²⁶⁶ These arguments are not strictly speaking arguments for compatibilism, that is, for the thesis that there *is* moral responsibility even in a deterministic world. It might be the case that we have pragmatic reasons to talk about moral responsibility, or that we cannot help but having a certain praxis, even if moral responsibility does not exist. The one truly compatibilist argument that Strawson advances is a metaethical one.

Strawson's metaethics is curious, however, and it is hard to say whether his concept of moral responsibility is expressivist or naturalist. When he discusses the praxis of holding people responsible, he seems to think that what we do when we blame or praise is to express an attitude towards the person praised or blamed. But he also seems to give a naturalist analysis of what it means to be morally responsible for an action. I think the best interpretation of Strawson's theory is that the speech act of praising or blaming has the expression of an attitude as an important component, while the fact that somebody is morally responsible is a natural fact. Moral responsibility depends, in Strawson's view, on the particular kind of interactions and relationships we can have with sane, adult humans. When we interact with other animals, little children and severely mentally ill people we try to figure out what to expect from their behaviour, and try to think of ways to best care for, handle or train them. Other sane adults we treat like equals, and our relationship with them is qualitatively different. An essential part of this kind of relationship is to hold each other morally responsible.²⁶⁷ It seems to be Strawson's view that a person is morally responsible for an action simply in case he *is* capable of partaking in normal adult relationships and interactions, and did what he did on purpose. This implies compatibilism, since whether the universe is deterministic or indeterministic, there *are* people capable of

²⁶⁵ McKenna and Russell, ed, 2008, pp 26 and 31

²⁶⁶ McKenna and Russell, ed, 2008, pp 28 and 31

having normal adult relationships and interactions, and these people sometimes mean to do what they did.

A common incompatibilist strategy is to focus on something the compatibilist agrees is an excusing condition, and then show that this condition generalises to all agents under determinism. If a compatibilist argues that a severely schizophrenic person cannot be responsible for what he does since he was not in control of himself, the incompatibilist will argue that in a deterministic world nobody is ultimately in control of himself. This strategy cannot be used against Strawson. He argues that the hopeless schizophrenic is not responsible because he cannot partake in normal, adult relationships.²⁶⁸ Determinism cannot mean that nobody can partake in normal, adult relationships.

As with Lenman, we may ask ourselves whether Strawson really discusses desert-entailing moral responsibility. Strawson brings up normal human relationships as an argument for the praxis of holding people responsible. If he means that these relationships have moral value, and that we ought to hold each other responsible in order to produce more of this value, it would not be desert-entailing moral responsibility, but rather a consequentialist concept. However, I do not think this is the best way to understand him. It seems to be his view that responsibility-holding is part of what *constitutes* rather than produces normal human relationships. We should not hold sane adult humans responsible for their actions because we can reach some further goal or produce some further value that way, but simply because they did what they did. If that is the right interpretation of Strawson's article, he does discuss desert-entailing moral responsibility. Compatibilism will still follow, since the moral responsibility of any agent will be a natural fact that can exist in both deterministic and indeterministic worlds.

²⁶⁷ *ibid* pp 25 and 27

²⁶⁸ *ibid* p 23-24

6.5. Realism

The word “realism” can be used to denote different families of theories. Sometimes “realism” is used to denote theories according to which there are moral truths as well as moral obligations that hold regardless of the agent’s motivation. If so, Kantian constructivism is a species of realism. In this chapter, however, I will consistently use “realism” to denote theories according to which moral truths are *mind-independent*.

There are no implications from this kind of realism to compatibilism. According to realism, Jimmy’s car theft was wrong if “stealing is wrong” is a mind-independent moral fact.²⁶⁹ Is this moral fact independent of whether the universe is deterministic? There is no obvious answer to this question. We can be certain that entailment relations, what principles it would be rational to accept or reject and the existence of people who are capable of normal human relationships do not depend on determinism. However, there is nothing obvious about mind-independent moral facts. It is not even obvious that such entities could exist at all, let alone that they could under determinism. Since there are no implications from moral realism to compatibilism, it is not surprising that we find both compatibilists and incompatibilists among realists.

Robert Kane argues that there is a connection between meta-normative realism and libertarian free will. He thinks that *ultimate* moral responsibility and moral agency in the *fullest sense* requires that agents have libertarian free will; that we are the ultimate originators of our actions, that the buck stops with us.²⁷⁰ Determinism, on Kane’s view, precludes “objective worth”.²⁷¹ This is what a person, thing or action has if it is not just thought to be valua-

²⁶⁹ It would, of course, also be wrong if “actions that fail to maximize utility are wrong” were a moral fact, or if “actions that manifest a dishonest character are wrong”. For the sake of argument I assume a deontological ethical theory to be true.

²⁷⁰ Kane, 1998, pp 4, 33-37, 97

²⁷¹ *ibid* p 97

ble by people, but valuable from a completely objective standpoint, from the view of the universe so to speak.²⁷² As pointed out at the beginning of this dissertation, a person can only have acted right or wrong if she acted freely. To act freely is to be the origin of one's actions and choose between different options. I have argued that we are the origin of our actions in the morally relevant sense if we are the origin from a practical perspective, if we must choose our actions and cannot delegate this choice back to the past and the laws of nature. Kane, however, thinks that the morally relevant view is the view of the universe, not that of a deliberating agent. This perspective is what shows us both what is truly valuable and who is truly free. If, from this perspective, we are not originators, not ultimate buckstoppers, then we are not free in the morally relevant sense, and we cannot have ultimate moral responsibility for our actions or moral agency in the fullest sense.^{273 274}

Kane thinks that determinism threatens much more than moral responsibility and moral agency. There could not be genuine creativity, autonomy, individuality, dignity or life-hopes if the world were deterministic.²⁷⁵ The key-word here is "genuine". Kane admits that there are compatibilist versions of all these concepts.²⁷⁶ He simply thinks that the compatibilist versions fall short of giving us everything worth wanting.²⁷⁷ That we cannot have *ultimate* moral responsibility or moral agency *in the fullest sense* if the world were deterministic is compatible with the possibility of some kind of compatibilist responsibility and moral agency under determinism. According to Kane's view, however, these concepts would be diluted compared to their incompatibilist counterparts.

²⁷² *ibid* pp 97-98

²⁷³ *ibid* p 97, quotation, and 97-98

²⁷⁴ The idea that some perceived "view of the universe" would only show us facts about what *is*, and nothing at all about what we ought to do, can also be found in Lenman, 2009, p 6. Nagel, 1986, pp 141-143 also discusses this claim, although he ultimately rejects it.

²⁷⁵ Kane, 1998, pp 81-88

²⁷⁶ *Ibid* pp 89-90

²⁷⁷ *Ibid* pp 91-97

I think Kane's argument has some force, because I understand what Nagel discusses in *the View from Nowhere*, what Pereboom calls "the problem of the disappearing agent" and what Bok calls "mechanism". When regarding the world from a theoretical, causal perspective, the actions of human beings do not look significantly different from other events in the world. Causal chains flow through us like they flow through anything else; there is no buck that ultimately stops with us agents. Of course human actions are different in *some* respect from other events; human actions are preceded (and caused, if epiphenomenalism is false) by certain mental events. The problem is that *some* difference could also be found between, for example, landslides and other events; landslides are, unlike other events, caused by certain geological and meteorological conditions. Just pointing out that those events we call human actions have some distinguishing features does not explain why they should be given special moral status, since the events we call landslides (or blizzards, or earthquakes, etc) also have distinguishing features. How could it then be that human actions can be *right* or *wrong*, when we think that other events are merely desirable or undesirable, fortunate or unfortunate? As Bok points out, the fact that humans are sometimes influenced by praise and blame is not a satisfying answer. Bok's computer example shows that it is not a sufficient condition for moral agency.²⁷⁸ I also think people could be moral agents even if they do not care about praise and blame, as long as they grasp the concepts of right and wrong, so it does not seem to me to be a necessary condition either.

There are people whom we may call natural compatibilists, who simply cannot see the problem that Kane, Nagel, Pereboom and Bok try to point out. To some people, it is simply *obvious* that the differences we can point to between human actions and other events, like human actions being preceded by intentions and thoughts, are morally relevant. I do not really know if the dispute between philosophers who think there is obviously no problem and

²⁷⁸ See Section 2.5.

those who think there is can be settled. For my own part, I think it is *prima facie* problematic that there does not seem to be any deep metaphysical difference between human actions on the one hand and other events on the other hand, that the buck does *not* stop with us, if human beings lack libertarian free will or the power of agent-causation.²⁷⁹ I also think that PDR shows that the *prima facie* problem is no real problem. A moral realist could, of course, agree with me here. Most moral realists believe there are intuitions that guide us to moral truth but that not all intuitions are truth-tracking. I have given several reasons to doubt our incompatibilist intuitions. A moral realist should therefore be wary of those and put greater trust in PDR.

Derek Parfit has his own argument for compatibilism about moral agency and determinism – an argument that fails if we assume, with Parfit, that metaethical theories according to which morality is mind-dependent are false. Parfit first distinguishes between a hypothetical and a categorical sense of “could have done otherwise”. “Could have done otherwise” in the categorical sense means that one could have done otherwise given exactly how the past and the laws of nature are. “Could have done otherwise” in a hypothetical sense means that one could have done otherwise if, for example, one had wanted to. Parfit writes that determinism threatens morality only if “alternative”, “can” and related words are interpreted in the categorical way. Parfit then argues that since moral thoughts can affect our future behaviour, a hypothetical interpretation of “can” makes most sense. *Moral deliberation* only requires that different alternatives are hypothetically open.

“Someone might now object:

²⁷⁹ Kane is an event-libertarian. However, I think I agree with Pereboom that although human actions would be significantly different from other events *even from a theoretical perspective* if we had a special power of causation that nothing else in nature has, a mere event-libertarian kind of freedom, some little quantum indeterminacy inserted here or there, would not mean that we were significantly different from anything else.

If all our decisions, choices, and acts are causally inevitable, we would have acted differently only if we had miraculously defied, or broken, the laws of nature. It is pointless to ask whether we ought to have acted in some way that would have required such a miracle.

Such questions, however, can be well worth asking. What we do often depends on our beliefs about what we ought to do. And if we come to believe that some act of ours was wrong, or irrational, because we ought to have acted differently, this belief may lead us to try to change ourselves, or our situation, so that we do not act wrongly, or irrationally, in this kind of way again. These changes in us or our situation may affect what we later do. It does not matter that, for us to have acted differently in the *past*, we would have had to perform some kind of miracle. If we come to believe that we ought to have acted differently, this change in our beliefs may cause it to be true that in similar cases, without any miracle, we *do* in the *future* act differently. That is enough to make it worth asking whether we ought to have acted differently.²⁸⁰

Parfit thinks this argument shows that right- and wrongdoing are compatible with determinism, although on his view desert-entailing moral responsibility is not. But the only thing he has actually argued for is that our *thoughts* about right and wrong can affect our behaviour, and no incompatibilist philosopher ever denied that. A moral agency incompatibilist like Haji would think error theory is true if determinism holds, since on his view “right”, “wrong” and “ought” cannot refer to anything if that is the case.²⁸¹ If physicists were to prove determinism tomorrow, Parfit could obviously not convince Haji that there is moral agency after all, by pointing out to him that people still *talk* and *think* in terms of right and wrong, and that these thoughts may affect their future actions. Parfit’s own arguments entail that this is beside the point, since later in his book Parfit argues passionately that

²⁸⁰ Parfit, 2011a, p 261

²⁸¹ Haji, 1999. I call him a moral agency incompatibilist since he is a compatibilist about moral responsibility and determinism, but incompatibilist about moral agency. Haji thus thinks that an agent can be morally responsible without being a moral agent.

moral thoughts and moral deliberation, no matter how perfect and idealised, do not constitute real moral reasons. Real moral reasons and moral facts are mind-independent.²⁸² ²⁸³Parfit cannot, on pain of giving up this metaethical view, suddenly claim that compatibilism about moral agency and determinism is true because moral deliberation exists even in a deterministic universe.

If Parfit had been, for instance, a constructivist of Korsgaard's kind, then he *would* have a good compatibilist argument here. Whether determinism is true or not, whether mechanism is true or not, we can base our actions on principles derived from our practical identity as rational animals. Nevertheless, if having a practical identity and principles is not enough, then the worries of Haji, Kane and other incompatibilists remain. These worries need not be unsolvable. As I wrote earlier, a moral realist could claim that PDR is true and that intuitions that clash with PDR fail to track the truth. PDR does not just point out, irrelevantly if you are a realist, that people could still talk and think in moral terms if determinism were true. PDR says that since morality is essentially action-guiding, factors that are irrelevant for a deliberating or advising agent are also irrelevant when judging the rightness or wrongness of an action, and a realist could agree. There are also other arguments in the free will debate that can be used by a moral realist. However, the simple kind of compatibilist argument that Parfit advances, pointing out that we could still use moral principles in deliberation if determinism were true, still thinking and speaking in moral terms, only works given a different metaethics than his own.

I have argued that Kantian constructivism, Strawsonian naturalism and Lenman's contractualism imply compatibilism, that Korsgaard's version of

²⁸² Parfit, 2011b, pp 275-288

²⁸³ Parfit also argues that the case that one ought to do something means that one has a reason to do so, and that one ought not to do something means that it is wrong Parfit, 2011a, pp 33 and 151. His oughts, wrongs and rights are thus as realist as his reasons are.

constructivism even implies something akin to PPC (since we can have practical freedom and public reasons and be rational animals regardless of our personal history or the causal chains behind our actions), and that Humean constructivism at least strongly suggests compatibilism. This may not be the case for all metaethical theories according to which the nature of morality is mind-dependent or naturalistic in one way or another. Still, a theory according to which “It was wrong of Jim to steal the car” can be analysed in terms of emotions, attitudes or ideas would seem *prima facie* to fit compatibilism, since emotions, attitudes and ideas certainly exist whether the universe is deterministic or indeterministic. Moral realism, while not implying incompatibilism, has the drawback (from a compatibilist viewpoint) that nobody can point to mind-independent moral facts or likewise moral responsibility and say “Look, this thing obviously exists even if determinism is true”, since there is nothing obvious about their existence in the first place.

Still, I have argued that classic incompatibilist arguments cannot be directed towards PPC without begging the question, that there are reasons not to put too much trust in intuitions elicited by far-fetched thought experiments, that the thought experiments traditionally used by incompatibilists or causal compatibilists do not lend as much support to these views as the proponents believe, and that there are good reasons to believe in PDR, which in turn supports PPC. These are good reasons for a realist to believe in PPC, even though realism itself does not imply either compatibilism or incompatibilism.

7. Conclusion

In this dissertation I have argued for a thesis I call practical perspective compatibilism, or PPC for short. PPC is the thesis that *practical freedom* is all the freedom we need in order to be moral agents and morally responsible for our actions. A moral agent is an agent who can do right or wrong. I claim that being a moral agent is a necessary condition for being morally responsible for one's actions, while being morally responsible is a sufficient condition for being a moral agent. In this claim I differ from Ishtiyaque Haji, who thinks people can be morally responsible although they are not moral agents. I find his view terribly counter-intuitive; I find it self-evident that moral agency is necessary for moral responsibility. The reason why moral responsibility is not necessary for moral agency is that some moral theories do not make use of this concept at all. If utilitarianism is true, for instance, there is no such thing as (desert-entailing) moral responsibility, but there could still be moral agents. I argue that two conditions need to be fulfilled for there to be moral responsibility in the world; firstly that there are moral agents, and secondly that the true or the best normative-ethical theory is one employing this concept.

Why assume that practical freedom is sufficient for moral agency and responsibility? Some philosophers have argued that it is not even sufficient for rational deliberation. The argument goes roughly like this: I cannot rationally deliberate about what I believe to be determined. If I believe everything to be determined, I cannot rationally deliberate about anything. All deliberation would therefore be irrational if determinism were true and we knew about it.

The argument for the claim that we cannot deliberate about what we believe to be determined is that we cannot deliberate about such things as whether the sun should rise tomorrow. However, the real reason we cannot deliberate about things like whether the sun should rise tomorrow is that our wills are inefficacious regarding tomorrow's sunrise. It is rational to deliberate about actions where my will is efficacious, meaning that I will do A if I decide to A and not-A if I decide to not-A.

Now, Kant and some Kantian philosophers believe that we need moral principles to deliberate. Agents who must choose what to do must have principles telling them what they ought to do, otherwise their decisions would be random and not real *choices* at all. Moral principles must have a certain form and a certain content to serve their functions as guides for a practically free agent, and an entire moral system can thus be inferred from the mere fact that there is an agent who must choose what to do. If the fact that agents have practical freedom entails facts about the rightness and wrongness of their actions, then practical freedom is obviously sufficient for moral agency, and PPC is true.

Although such inferences of morality from freedom have been popular among Kantians, most non-Kantians have been unconvinced. Presenting an unassailable inference with no questionable premises has yet to be done. However, there is another way to argue for PPC. It is widely agreed that morality is action-guiding. This means that morality is something we should use in deliberation and advice. That an action is morally right speaks in favour of it and the other way around. The principle of deliberative relevance, PDR, says that only factors that are relevant for a deliberating or advising agent can be relevant when judging whether somebody did right or wrong. Although PDR does not follow logically from the fact that morality is action-guiding and rightness and wrongness are concepts to be employed in deliberation and advice, I do think it is made highly plausible by considering the action-guiding nature of morality, right and wrong. The question now is

what it means for a factor to be relevant for a deliberator or adviser. I argue that it can be relevant in two ways; either by making rational deliberation and advice possible in the first place, or by figuring as a reason pro or con doing an action. Factors like the consequences of an action, that it was done from a virtuous motive, that it followed the categorical imperative, and so on, are relevant because they can figure as reasons for doing an action. That an agent has practical freedom is relevant because it makes rational deliberation and advice possible in the first place. However, indeterminism, determinism, agent-causation, the personal history of an agent and so on are not relevant for deliberators and advisers, and they are thus not relevant when judging whether an action was right or wrong. Therefore, practical freedom is sufficient, freedom-wise, for moral agency. A moral agent might also need, for instance, the capacity to relate to others in the right way. If so, it is because this capacity is necessary for being able to deliberate about *moral* matters. However, determinism, indeterminism, agent-causation or the right personal history is not necessary for any kind of deliberation, and is thus morally irrelevant according to PDR.

That an action was right or wrong is not only a judgement we make when deliberating or advising but also a judgement we might make from a third-person standpoint. I argue, however, that a third-person judgement must be able to function as a hypothetical advice. The property of being right or wrong cannot change depending on whether we advise or merely judge. Therefore, PDR holds for third-person judgements as well.

Thus far, the argument has been for the sufficiency of practical freedom for moral agency. I also argue that we should regard moral responsibility as connected to right- and wrongness. If we assume that the true or best ethical theory employs the concept of moral responsibility, we should consider people praiseworthy when they do what is right, subjectively right, or seemed right to them, and the same with blame and wrongness. This is *prima facie* more plausible than regarding moral responsibility as completely discon-

nected from rightness and wrongness. In real life, we do tend to blame people because they did wrong and praise them because they did right, and we praise and blame from within the same practical perspective as we have when deliberating or advising. We praise and blame by citing reasons for or against the action performed, or by giving a kind of late advice (“you should not have done that!”). I argue that if no convincing arguments for the incompatibilism of moral responsibility and determinism (or indeterminism) appear, we should believe that the same conditions hold for moral responsibility as for moral agency.

There are many people whom we consider less than fully free in the morally relevant sense, for example, neurotics, addicts, people in Milgram scenarios and people from a certain background who unquestioningly follow in their parents’ footsteps. If practical freedom were an all-or-nothing affair, this would be a problem for PPC. Either I would have to argue that these people are fully free in the morally relevant sense, or I would have to argue that they lack practical freedom and do not choose what to do. However, I define practical freedom in a way that allows for degrees. As soon as one must choose what to do and believe one’s will to be efficacious, one has practical freedom. In order to have maximal practical freedom, one must really have an efficacious will, have full information about one’s considered options, and be able to do what one tries to do when choosing an option: that is, picking the option that is best according to one’s own values. This explains why people in a Milgram scenario or people from certain backgrounds often have less than full practical freedom. When it comes to some cases of neurotics and addicts they might be as free as anybody, although their options look different. What is easy for other people may be very difficult for them, and this should be taken into account when considering praise- and blameworthiness.

PPC is more radical than many other compatibilist theories, since a very thin kind of freedom is sufficient for having at least some moral agency and moral responsibility. The sheer radicalism of PPC might be considered a drawback. However, PPC has some distinct advantages over other kinds of compatibilism, according to which one can be a moral agent and morally responsible if one's actions ensue from the right kind of causal chain. Firstly PPC relies on PDR, while most other compatibilist theories rely on intuitions elicited by more or less fanciful thought experiments. Now there is probably no getting around the fact that we ultimately have to rely on intuitions when doing philosophy, but we have reason to be wary of intuitions elicited by far-fetched scenarios that do not resemble anything we have encountered in real life. PDR is more trustworthy. This is a reason to believe that PPC, rather than some less radical theory about freedom and moral agency/responsibility, is the true one. Secondly, PPC means that moral agency and responsibility are compatible not only with determinism as well as indeterminism, but also with epiphenomenalism, while many other compatibilist theories require that some particular theory of mental causation be true. PDR claims that anything irrelevant to a deliberating and advising agent is also morally irrelevant. Now whether the mental life is an epiphenomenon or not I can rationally deliberate as long as my will is efficacious, and the mental life being epiphenomenal is clearly something that cannot give me reason to do one thing rather than another. It is thus morally irrelevant. I could still have my practical freedom. It could still be the case that my will is efficacious in that I do A if I decide to A and not-A if I decide to not-A. As long as this condition holds, I can still be a moral agent and morally responsible. This in itself is of course no reason to believe in PPC, but a nice consequence of it being true. We do not have to solve complex problems in the philosophy of mind in order to know that actions can be right or wrong.

PPC might seem very counterintuitive if practical freedom merely was the mistaken belief that one is free. That a mistaken belief would somehow grant one moral agency and responsibility would be a very strange claim. Practical freedom might seem like a mistaken belief since it is freedom from the perspective of a deliberating or advising agent – if this “freedom” evaporates when we regard things from a theoretical, scientific freedom, was it not simply an illusion? It would be, if practical freedom just meant that *it does not seem like* my behaviour is determined or subject to randomness from a practical perspective. However, it is instead the case that determination as well as quantum randomness or agent-causation are *irrelevant* from a practical perspective, since they cannot help me decide what I should do. Practical freedom thus has nothing to do with being mistaken about one’s situation.

PDR is an argument for PPC. Even so, there are also some well-known counterarguments both against compatibilism in general, and against so-called time-slice compatibilist theories (of which PPC is a species). Galen Strawson’s basic argument, van Inwagen’s consequence argument and various thought experiments target groups of theories where PPC is included. These arguments aim to show that we cannot be morally responsible if the world is deterministic, or if it is indeterministic, or if our values do not have the right kind of history. They may be effective against certain kinds of compatibilism, but not against PPC. Against PPC they are question-begging, since they simply assume that PDR is false, and that one will arrive at trustworthy moral intuitions by focusing on factors that are irrelevant according to PDR.

Still, one might think that thought experiments would have some force if they showed that some terribly counter-intuitive implications would follow from PPC. If that were the case, we might be well-advised to abandon PPC as well as PDR. However, this is not the case. By reflecting on the thought experiments used by incompatibilists or causal compatibilists we can see that

it is not counter-intuitive at all to claim that the main characters of their examples are morally responsible for what they do. It might seem like PPC implies that a person after manipulation could be morally responsible for what he did before the manipulation, and this might seem counter-intuitive after all. However, this does not follow. I argue that a moral agent can only be morally responsible for acts that the same moral agent has performed. If A did not choose the actions of B, A and B are not the same moral agent, and B cannot be morally responsible for what A did.

One might also believe that PPC cannot account for the mitigating effect a bad childhood can have on judgements of moral responsibility. A bad childhood is normally no reason to choose one option over another, and it does not make deliberation or advice either possible or impossible, and so it should be considered morally irrelevant according to PDR. However, even if a bad childhood does not directly affect moral responsibility, it can do so indirectly. People with bad childhoods might not deliberate about options that would be better according to their own values than what they actually choose, and thus they have their practical freedom diminished. Their options may also look different from the options of other people, since what comes easy for others may be difficult for them, and how difficult something was to perform affects how praise- or blameworthy one was for it.

Thus far, I have shown both that there is a strong *prima facie* argument in favour of PPC, namely PDR, and that arguments against compatibilism or against time-slice compatibilism fail if directed against PPC. I can thus conclude already that we have good reasons to believe in PPC. If we move to the metaethical level, PPC gains further support. Christine Korsgaard's metaethical theory implies that a compatibilist theory along the lines of PPC must be true. Peter Strawson's naturalism, James Lenman's contractualism and Kantian constructivism in general imply that compatibilism is true. Humean constructivism at least strongly suggests compatibilism. If a

metaethical theory says that rightness, wrongness and moral responsibility are constituted by something that obviously exists regardless of determinism or indeterminism, then this theory will imply compatibilism. On the other hand, there is no metaethical theory that implies incompatibilism. No metaethical theory claims that rightness, wrongness or moral responsibility consists in something that obviously *cannot* exist if the universe is deterministic.

According to moral realism, an action is right if it corresponds to a mind-independent moral fact saying it is right. No definite argument can be given either for the view that such facts could exist under determinism or for the view that they could not. If one is a moral realist, there is thus no implication from one's metaethical view to a certain stance in the free will debate. I argue that a moral realist still has reason to believe in PPC; PDR applies as much to moral realists as to others. For some metaethical views, however, PDR actually becomes superfluous for proving compatibilism. When these views are supposed, one can infer compatibilism from the metaethical theory.

My main conclusion is thus that there are good reasons to believe in practical perspective compatibilism. This means that our actions can be right or wrong regardless of whether the universe is deterministic or indeterministic. It also means that the question of whether people can be morally responsible comes to rest on the level of normative ethics. If the best normative-ethical theory employs this concept, people can; otherwise they cannot. but for reasons that have got nothing to do with determinism or indeterminism.

Bibliography

- Aristotle. 1998 (third edition). *The Nicomachean Ethics*. Translation to English by David Ross. Oxford: Oxford University Press.
- Björnsson, Gunnar and Persson, Karl. 2011. "The Explanatory Component of Moral Responsibility". *Noûs*. 00:0 1-29
- Bok, Hilary. 1998. *Freedom and Responsibility*. Princeton: Princeton University Press.
- Campbell, Joseph Keim et al ed. 2004. *Freedom and Determinism*. Cambridge, Massachusetts: The MIT Press.
- Carlson, Erik. 2002. "Deliberation, Foreknowledge, and Morality as a Guide to Action." *Erkenntnis* 57: 71-89
- Chisholm, Roderick. 1976. *Person and Object*. London: George Allen and Unwin.
- Darwall, Gibbard and Railton, 1992 "Toward *Fin de siècle* Ethics: Some Trends." *Philosophical Review*. No 101.
- Davidson, Donald. 2001. *Essays on Actions and Events*. Oxford: Clarendon Press.
- Denett, Daniel, ed. 1981. *Brainstorms, Philosophical Essays on Mind and Psychology*. Brighton: Harvester Press.
- Enoch, David. 2011. *Taking Morality Seriously – a Defence of Robust Realism*. Oxford: Oxford University Press.

- Feinberg, Joel. 1980. "The Interest in Liberty on the Scales". *Rights, Justice, and the Bounds of Liberty: Essays in Social Philosophy*. Princeton: Princeton University Press.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell Publishers Ltd.
- Fischer, John Martin and M Ravizza. 2000. "Replies". *Philosophy and phenomenological research* 61: 467-469.
- Fischer, John Martin. 2010. "The Frankfurt Cases: The Moral of the Stories." *Philosophical Review* 119 (3):315-336.
- Foster, L and Swanson, J. 1970. *Experience and Theory*. Massachusetts: Universities of Massachusetts Press.
- Frankfurt, Harry G. 1969. "Alternate Possibilities and Moral Responsibility". *The Journal of Philosophy*. Vol 66, No 23. 829-839.
- Ginet, Carl. 1990. *On Action*. Cambridge: Cambridge University Press.
- Glatz, Richard M. 2007. "The (near) necessity of alternate possibilities for moral responsibility." *Philosophical Studies* 139 (2):257 - 272.
- Haji, Ishtiyaque. 1999. "Moral Anchors and Control". *Canadian Journal of Philosophy* 29: 175-204
- Haji, Ishtiyaque. 2009. "Freedom and Practical Reason". *Ethic Theory Moral Prac* 12: 169-179
- Hanson, Sven Ove. 2010. "Objective or Subjective 'Ought'?" *Utilitas*. Vol 22 No 1.
- Hawthorne, John. 2001. "Freedom in Context". *Philosophical Studies*. Vol 104, No 1.
- Hobart, R. E. 1934 "Free Will as Involving Determination and Inconceivable Without it." *Mind*. no 169.

- Honderich, Ted. 2002. *How Free Are You?* Oxford: Oxford University Press.
- Honderich, Ted. 2005. *On Determinism and Freedom*. Edinburgh: Edinburgh University Press.
- Kane, Robert. 1985. *Free Will and Values*. Albany: State University of New York.
- Kane, Robert. 1996. *Through the Moral Maze*. New York: North Castle Books
- Kane, Robert. 1998. *The Significance of Free will*. Oxford: Oxford University Press.
- Kane, Robert. Editor. 2002. *The Oxford Handbook of Free Will*. Oxford: Oxford University Press.
- Kane, Robert. Editor. 2011. *The Oxford Handbook of Free Will*. Second Edition. New York: Oxford University Press.
- Kant, Immanuel. 1997. *Grundläggning av Sedernas Metafysik*. Translation to Swedish by Joachim Retzlaff. Uddevalla: Bokförlaget Daidalos AB.
- Kant, Immanuel. 2001. *Kritik av det Praktiska Förnuftet*. Translation to Swedish by Agne Lundquist. Lund: Bokförlaget Argos.
- Kant, Immanuel. 2004. *Kritik av det Rena Förnuftet*. Translation to Swedish by Jeanette Emt. Stockholm: Thales.
- Kapitan, Tomis. 1986. Deliberation and the Presumption of Open Alternatives. *The Philosophical Quarterly* vol. 36 No. 14 230-251
- Kaye, Anders. 2007 "The Secret Politics of the Compatibilist Criminal Law." *Kansas Law Review*. vol 55
- Kim, Jaegwon. 2006. *Philosophy of mind*. Second edition. Cambridge: Westview Press.

- Kierkegaard, Sören. 1968. Sörken Kierkegaards Papirer. *Anden förögede Udgave ved Niels Thulstrup*. Gyldendal: Köpenhamn.
- Kierkegaard, Sören. 2002a. *Antingen-Eller. Ett livsfragment. Första delen*. Translation to Swedish by Stefan Borg. Önnköping: Nimrod
- Kierkegaard, Sören. 2002b. *Antingen-Eller. Ett livsfragment. Andra delen*. Translation to Swedish by Stefan Borg. Önnköping: Nimrod
- Korsgaard, Christine. 1996a. *the Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, Christine. 1996b. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Korsgaard, Christine. 2008. *The Constitution of Agency*. Oxford: Oxford University Press.
- Lenman, James. 2006. "Compatibilism and Contractualism: The Possibility of Moral Responsibility." *Ethics*, no 117.
- Lenman, James. 2009. "Naturalism Without Tears". *Ratio* no 22 (1) 1-18
- Lewis, David. 1996. "Elusive Knowledge". *Australasian Journal of Philosophy*. Vol 74, No 4; December.
- McKenna, Michael. 2008. "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument". *Philosophy and Phenomenological Research*. Vol. LXXVII No 1
- McKenna, Michael and Russell, Paul, ed. 2008. *Free Will and Reactive Attitudes*. Surrey: Ashgate Publishing Ltd.
- McLeod, Owen. "Desert". <http://plato.stanford.edu/entries/desert/> First published Tue May 14, 2002; substantive revision Wed Nov 12, 2008
- Mele, Alfred. 2001. *Autonomous Agents. From Self-Control to Autonomy*. Oxford: Oxford University Press.

- Mele, Alfred. 2006. *Free will and luck*. Oxford: Oxford University Press.
- Mele, Alfred. 2009. *Effective Intentions*. Oxford: Oxford University Press.
- Milgram, Stanley. 1975. *Lydnad och Auktoritet*. Stockholm: Wahlström och Widstrand.
- Myers, David G. 2002, 7th Edition. *Social Psychology*. New York: McGraw Hill.
- Nagel, Tomas. 1986. *The view from nowhere*. New York: Oxford University Press.
- Nahmias, Coates and and Kvaran. 2007. "Free Will, Moral Responsibility and Mechanism; Experiments on Folk Intuitions." *Midwest Studies in Philosophy*. XXXI
- Nelkin, Dana. 2009. "Two Standpoints and the Belief in Freedom." *The Journal of Philosophy* vol. 96 No. 10
- Nelkin, Dana. 2007. "Do We Have a Coherent Set of Intuitions about Moral Responsibility?" *Midwest Studies in Philosophy*. Vol 31, issue 1, pp 243-259
- Nichols, Shaun, and Knobe, Joshua. 2007. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." *Noûs* 41 (4):663–685
- O'Connor, Timothy. 2002. *Persons and Causes. The Metaphysics of Free Will*. Electronic resource at <http://www.sub.su.se>
- Otsuka, Michael. "Incompatibilism and the Avoidability of Blame". *Ethics*, 108, July 1998, pp 685-710
- Parfit, Derek. 2011a. *On what matters. Volume one*. Oxford: Oxford University Press.

- Parfit, Derek. 2011b. *On what matters. Volume two*. Oxford: Oxford University Press.
- Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Pereboom, Derk. 2008 (the version I printed out, 13/6 -08.). A Compatibilist Account of the Beliefs Required for Rational Deliberation. http://experimentalphilosophy.typepad.com/2nd_annual_online_philosophy/files/derk_pereboom.pdf
- Rieber, Steven. 2006. "Free Will and Contextualism" *Philosophical Studies*. 129: 223-252.
- Ryseview, Patrick. 2011. <http://plato.stanford.edu/entries/contextualism-epistemology/> First published Fri Sep 7, 2007; substantive revision Mon Nov 14, 2011.
- Sher, George. 2001. "Blame for Traits". *Noûs*. 35:1.
- Skinner, Burrhus Frederic. 1948. *Walden Two*. New York: Macmillan, cop.
- Smart, John Jamieson Carswell. 1961. "Freewill, Praise and Blame." *Mind*. 70. 291-306.
- Smith, Matthew Noa. "Practical Imagination and Its Limits." *Philosophers Imprint* 3 pp 1-19
- Street, Sharon. 2010. "What is Constructivism in Ethics and Metaethics?". *Philosophy Compass* 5/5,
- Wallace, R Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- van Inwagen, Peter. 1980. "The incompatibility of responsibility and determinism". *Studies in Applied Philosophy*. Vol 2.

- van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- van Inwagen, Peter. 2000. "Free Will Remains a Mystery". *Philosophical Perspectives*. No 14.
- van Inwagen, Peter. Forthcoming in *Ethics*: "How to Think about the Problem of Free Will" 2011-10-03 it can be found at <http://philpapers.org/rec/VANHTT>
- Vargas, Manuel. Philosophy and the Folk: 2006. "On Some Implications of Experimental Work For Philosophical Debates on Free Will." *Journal of Cognition and Culture*. Vol. 6, No. 1-2
- Wallace, R Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- Watson, Gary. 2004. *Agency and Answerability*. Oxford: Clarendon Press.
- Wedberg, Anders. 1966. *Filosofins Historia: Från Bolzano till Wittgenstein*. Stockholm: Thales.
- Widerker, David. 1995. "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities". *The Philosophical Review*. Vol. 104, No. 2.
- Wilkes, Kathleen 1988. *Real People: Personal Identity Without Thought Experiments* Oxford: Oxford University Press.
- Williams, Michael. 2001 *Problems of Knowledge. A Critical Introduction to Epistemology*. Oxford: Oxford University Press.
- Willaschek, Marcus. 2009. "Non-Relativist Contextualism about Free Will." *European Journal of Philosophy*. Volume 18, issue 4, pp 567-587.

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

Zaibert, Leo. 2009 "The Paradox of Forgiveness." *Journal of Moral Philosophy*. no 6.

Stockholm Studies in Philosophy

Published by Stockholm University Editors: Gunnar Svensson and Torbjörn Tännsjö

1. Stig Kanger. *Provability in Logic*. Stockholm 1957. 47 pp.
2. Jan Berg. *Bolzano's Logic*. Stockholm 1962. 214 pp.
3. Dag Prawitz. *Natural Deduction. A Proof-Theoretical Study*. Stockholm 1965. 113 pp.
4. Lars Bergström. *The Alternatives and Consequences of Actions. An essay on some fundamental notions in teleological ethics*. Stockholm 1966. 141 pp.
5. Torbjörn Tännsjö. *The Relevance of Metaethics to Ethics*. Stockholm 1976. 226 pp.
6. Lars O. Ericsson. *Justice in the Distribution of Economic Resources. A critical and normative study*. Stockholm 1976. 150 pp.
7. Per-Erik Malmnäs. *From Qualitative to Quantitative Probability*. Stockholm 1981. 73 pp.
8. Gunnar Svensson. *On Doubting the Reality of Reality. Moore and Wittgenstein on sceptical doubts*. Stockholm 1981. 120 pp.
9. Torkel Franzén. *Provability and Truth*. Stockholm 1987. 81 pp.

10. Fredrik Stjernberg. *The Public Nature of Meaning*. Stockholm 1991. 173 pp.
11. Lars-Göran Johansson. *Understanding Quantum Mechanics. A realist interpretation without hidden variables*. Stockholm 1992. 177 pp.
12. Claudio Marcello Tamburrini. *Crime and Punishment?* Stockholm 1992. 293 pp.
13. Markku Leppäkoski. *The Transcendental How. Kant's transcendental deduction of objective cognition*. Stockholm 1993. 280 pp.
14. Folke Tersman. *Reflective Equilibrium. An Essay in Moral Epistemology*. Stockholm 1993. 144 pp.
15. Hans Ruin. *Enigmatic Origins. Tracing the Theme of Historicity through Heidegger's Works*. 1994. 304 pp.
16. Björn Eriksson. *Heavy Duty. On the Demands of Consequentialism*. Stockholm 1994. 210 pp.
17. Cesare Cozzo. *Meaning and Argument. A theory of meaning centred on immediate argumental role*. Stockholm 1994. 216 pp.
18. Sören Häggqvist. *Thought Experiments in Philosophy*. Stockholm 1996. 206 pp.
19. Henrik Bohlin. *Groundless Knowledge. A Humean Solution to the Problem of Skepticism*. Stockholm 1997. 231 pp.

20. Daniel Birnbaum. *The Hospitality of Presence. Problems of Otherness in Husserl's Phenomenology*. Stockholm 1998. 200 pp.
21. Henrik Hållsten. *Explanation and Deduction. A Defence of Deductive Chauvinism*. Stockholm 2001. 165 pp.
22. Filip Widebäck. *Identity of Proofs*. Stockholm 2001. 85 pp.
23. Ulrik Kihlbom. *Ethical Particularism. An Essay on Moral Reasons*. Stockholm 2002. 159 pp.
24. Charlotta Weigelt. *The Logic of Life. Heidegger's Retrieval of Aristotle's Concept of Logos*. Stockholm 2002. 236 pp.
25. Håkan Salwén. *Hume's Law. An Essay on Moral Reasoning*. Stockholm 2003. 154 pp.
26. Marcel Quarfood. *Transcendental Idealism and the Organism. Essays on Kant*. Stockholm 2004. 221 pp.
27. Jens Johansson. *Mortal Beings. On the Metaphysics and Value of Death*. Stockholm 2005. 180 pp.
28. Robert Callergård. *An Essay on Thomas Reid's Philosophy of Science*. Stockholm 2006. 164 pp.
29. Arvid Båve. *Deflationism: A Use-Theoretic Analysis of the Truth-Predicate*. Stockholm 2006. 207 pp.
30. Anna-Karin Andersson. *Libertarianism and Potential Agents. A Libertarian View of the Moral Rights of Foetuses and Children*. Stockholm 2007. 122 pp.
31. Karl Weigelt. *The Signified World. The Problem of Occasionality in Husserl's Phenomenology of Meaning*. Stockholm 2008. 343 pp.

32. Karl Karlander. *The Normativity of Thought and Meaning*. Stockholm 2008. 131 pp.
33. B. Ingemar B. Lindahl. *On Causal Attribution*. Stockholm 2009. 369 pp.
34. Nicholas Smith. *Towards a Phenomenology of Repression – a Husserlian Reply to the Freudian Challenge*. Stockholm 2010. 336 pp.
35. Sofia Jeppsson. *Practical Perspective Compatibilism*. Stockholm 2012. 213 pp.

Subscriptions to the series and orders for single volumes should be addressed to any international bookseller or directly to the distributor: eddy.se ab, P.O. Box 1310, SE-621 24 Visby, Sweden. Phone: +46 498 253900

Fax: +46 498 249789 E-mail: order@bokorder.se
<http://acta.bokorder.se>

ACTA UNIVERSITATIS STOCKHOLMIENSIS

Corpus Troporum Romanica Stockholmiensia

Stockholm Cinema Studies

Stockholm Economic Studies. Pamphlet Series

Stockholm Oriental Studies

Stockholm Slavic Studies

Stockholm Studies in Baltic Languages

Stockholm Studies in Classical Archaeology

Stockholm Studies in Comparative Religion

Stockholm Studies in Economic History

Stockholm Studies in Educational Psychology

Stockholm Studies in English

Stockholm Studies in Ethnology

Stockholm Studies in Film History

Stockholm Studies in History
Stockholm Studies in History of Art
Stockholm Studies in History of Ideas
Stockholm Studies in History of Literature
Stockholm Studies in Human Geography
Stockholm Studies in Linguistics
Stockholm Studies in Modern Philology. N.S.
Stockholm Studies in Musicology
Stockholm Studies in Philosophy
Stockholm Studies in Psychology
Stockholm Studies in Russian Literature
Stockholm Studies in Scandinavian Philology. N.S.
Stockholm Studies in Social Anthropology. N.S.
Stockholm Studies in Sociology. N.S.
Stockholm Studies in Statistics
Stockholm Theatre Studies
Stockholmer Germanistische Forschungen
Studia Fennica Stockholmiensia
Studia Graeca Stockholmiensia. Series Graeca
Studia Graeca Stockholmiensia. Series Neohellenica
Studia Juridica Stockholmiensia
Studia Latina Stockholmiensia
Studies in North-European Archaeology