Kassu Ian Kasperi Partanen

On Free Will and Moral Responsibility

A Defense of Hard Incompatibilism

NTNU
Norwegian University of
Science and Technology

Kassu Ian Kasperi Partanen

On Free Will and Moral Responsibility

A Defense of Hard Incompatibilism

NTNU

Norwegian University of
Science and Technology

Abstract

Skeptical views on free will are broadly deemed unattractive, and only a minority of philosophers hold the skeptical position that we lack free will. In this thesis, I will defend a hard incompatibilist account of free will. I will show that this skeptical position is plausible, and far less unattractive than many philosophers have claimed. I will do so by examining and critiquing the more influential positions of libertarianism and compatibilism. I argue that these accounts are unable to provide plausible answers to the question of how we can be said to have free will and be morally responsible. Then, I will argue why hard incompatibilism does not face the same difficulties as these positions and why I consider this account a more plausible alternative. In light of these findings, I will also argue that hard incompatibilism is not a threat to our interpersonal relationships, to our deliberation or our decision-making, or to our sense of morality and justice. The goal of these sections is to showcase why hard incompatibilism is a more attractive account than has been claimed and that adopting this view of free will need not, I argue, have undesirable personal and societal consequences.

Acknowledgements

First, I want to extend my respect to my elders by thanking my mom and dad for their continued support during the writing of this thesis. I also want to thank Anders Nes for his enthusiasm for this thesis and for his competent supervision. Lastly, I am grateful to my friends and fellow students for their support, valuable insights, and helpfulness during these years of research.

Table of contents

Introduction

The question of whether we have free will has been a central question throughout the history of philosophy. Many of the most influential philosophers throughout history, including Plato, Aristotle, Augustine of Hippo, Thomas Aquinas, Rene Descartes and Immanuel Kant, have taken up the topic (O'Connor, Franklin, 2022, foreword). Today, free will is still a much-discussed topic in philosophical literature. This is due to the timeless difficulties this topic presents us with, as well as modern scientific advancements that have a bearing on the philosophical debate. Despite the discourse on free will spanning millennia, philosophers are still torn about whether we have free will, and if we do, in virtue of what we possess this power. It is also an important topic, as whether we have free will and whether we believe we have free will has bearings on our views of ourselves and others, blame and praise, agency, and justice. Therefore, in dealing with the question of whether we have free will, this thesis concerns an essential question about how we ought to conduct our lives.

To some, it might seem absurd to claim anything but that we have free will. After all, it seems to each one of us like we have the freedom to do as we like, and that we freely deliberate and choose between options all the time. Others might claim the opposite, given that we are a product of nature and nurture, neither of which we had any freedom over. They might argue that we could not have become anything but what we became because of this and that we cannot be said to have free will as a result. Regardless of how one intuitively feels about this topic, I am going to show that this question is significantly more complex than our intuitions might tell us, but that progress can nevertheless be made in answering the question of whether we have free will.

I will begin by defining and explaining some key terms for this thesis, and the free will discourse at large. The most crucial term in this thesis is that of "free will" itself. There is not a universal agreement on what free will means.[1] However, a key aspect of what free will entails that most philosophers can agree upon is that it entails an ability to make decisions that make us morally responsible.

---

[1] Peter van Inwagen points out that some philosophers have drawn a distinction between libertarian and compatibilist forms of free will, and that he finds this distinction the worst among the confused ideas and bad terminologies in the free will debate (van Inwagen, 2008, p327).

Much of the tradition has taken "free will" to be a kind of power or ability to make decisions of the sort for which one can be morally responsible, but philosophers have also sometimes thought that free will might be required for a range of other things, including moral value, originality, and self-governance. (Kane, Fischer, Pereboom, Vargas, 2007, p1)

In other words, free will enables us to be held morally responsible, either due to us being able to do otherwise, being the first cause of an action, or both. While philosophers disagree on the specifics, free will concerns some form of freedom and ability that makes us morally responsible for exercising this freedom and ability, either constantly or intermittently. A relatively uncontroversial example of a free action of this kind is that of a grown, rational and coherent person stealing money out of greed while they could have abstained from doing so.

The above quote shows how closely tied the concept of moral responsibility it is to the concept of free will. While moral responsibility is not a universally agreed upon term, Derk Pereboom's characterization of the sense in which "moral responsibility" has been used in the philosophical debate is as follows:

..for an agent to be morally responsible for an action is for it to belong to her in such a way that she would deserve blame if she understood that it was morally wrong, and she would deserve credit or perhaps praise if she understood that it was morally exemplary. (Pereboom, 2007, p86)

If free will means an ability to make decisions for which we can be held morally responsible, moral responsibility would follow from having free will, at least for actions that were carried out with free will. It can seem difficult to see how moral responsibility would follow from anything but actions that were freely willed. After all, why would someone deserve praise or blame for something they did not freely will to do? Therefore, having moral responsibility seemingly then also entails that we acted with free will, and free will or moral responsibility seems to follow from the other holding true. Works on free will and moral responsibility also often address both concepts at once and tend to state or heavily imply that one entails the other. I also believe this is the case and I will therefore discuss free will and moral responsibility under the assumption that one follows from the other in this thesis. Some philosophers draw a stronger distinction between

these two concepts.[2] I will not discuss such accounts in detail as I find them inherently implausible and because discissions on this topic would be far removed from the other areas discussed in this thesis.

Another key concept discussed throughout this thesis is "determinism". As I will discuss a range of views, I will be relying on this simple definition of the concept that Robert Kane and colleagues have laid out:

> For present purposes, we can treat determinism as the thesis that at any time (at least right up to the very end) the universe has exactly one physically possible future. Something is deterministic if it has only one physically possible outcome. (Kane, Fischer, Vargas, Pereboom, 2007, p2)

In other words, if determinism holds true, things will necessarily play out in exactly one way, and no other outcome is possible. Determinism therefore poses us with an important question. Namely, if only one physically possible outcome is possible, how can we be said to have free will? If things could only play out one way, it would seemingly follow that we cannot impact the course of events, and if we cannot impact the course of events, we also seemingly do not have free will. While determinism is a more complex concept than laid out here, this is the core problem the concept poses for the free will debate.[3] Accounts claiming we have free will must therefore either deny determinism or show how determinism is compatible with free will.

With these concepts defined, I will proceed by laying out some common views on free will that are crucial for understanding the accounts discussed in this thesis. To understand what these views entail, we first need to understand the difference between incompatibilist and compatibilist accounts of free will. Incompatibilists hold that free will and determinism are mutually exclusive and that we can only act with free will if determinism does not hold true (Clarke, Capes, Swenson, 2021, foreword). Compatibilism, conversely, is the thesis that free will and

---

[2] An example of this is John Martin Fischer's account of Semicompatibilism. According to this view, determinism is compatible with moral responsibility but not compatible with free will (Fischer, 2007, 2012).

[3] Carl Hoefer writes that the italicized words need further explanation and investigation when providing the following definition of determinism: Determinism is true of the *world* if and only if, given a specified *way things are at a time t*, the way things go *thereafter* is *fixed* as a matter of *natural law* (Hoefer, 2023, 1*)*. However, the definition provided by Kane and colleagues provides a sufficient definition of determinism for the discussions of my paper as it encapsulates the core idea of determinism and the threats it poses to free will and moral responsibility.

determinism are compatible (McKenna, Coates, 2024, foreword). Incompatibilists who hold that we have free will are called libertarians (Kane, Fischer, Pereboom, Vargas, p3). Incompatibilists who hold that we do not have free will have in the contemporary debate been called hard incompatibilists (ibid).[4]

This thesis aims to argue that hard incompatibilism is a plausible account and that endorsing this view is a natural response when the weaknesses of other common accounts are highlighted. My methodology for doing so is first to lay out what common accounts of free will claim, why one might endorse them, and why I ultimately do not find these accounts convincing. My goal with this is to argue in favor of hard incompatibilism by highlighting the weaknesses of common alternatives and holistically show why I find hard incompatibilism to be the most plausible view available. I also aim to show that hard incompatibilism is not an unattractive account to endorse. Put another way, in this thesis, I argue that we do not have free will and that we do not have moral responsibility, and I also make the point that the undesirability of skeptical positions is exaggerated.

I will be discussing a wide range of views, all of which have been written extensively about. Therefore, most of the subsections of this thesis could easily warrant a thesis of their own. For this reason, I will concisely outline the respective views I discuss in this paper and focus on what I consider to be the most critical arguments in favor of and against them. I will focus on why incompatibilism holds true, why I do not believe we have free will, and the merits of hard incompatibilism. This means that many interesting arguments relevant to the topic will not be covered. Many of these topics also have a rich history, which I can only briefly outline. However, I believe that my discussion on libertarianism and compatibilism serves as a sufficient reason for accepting hard incompatibilism as a more plausible and possibly more desirable alternative. This is because I've focused on central and fundamental problems with libertarianism and compatibilism. Therefore, even if libertarian and compatibilist accounts have been defended in other ways, my arguments are intended to highlight fundamental weaknesses with these views that any reasonable defense of these positions cannot leave unaddressed.

---

[4] In the past, skeptics generally adopted a hard determinist view. However, few contemporary philosophers are hard determinists due to modern scientific discoveries. This point is discussed further in chapter 4.1.

The rest of the thesis will unfold as follows. In chapter 1, the subject matter is compatibilism. I will begin by briefly outlining the view and its nuances and presenting a well-known defense of the view by Harry G. Frankfurt (Frankfurt, 1969). I will then divert my focus to how compatibilists have attempted to argue against the consequence argument, as this argument is a key reason I endorse incompatibilism.[5] This discussion will partly focus on the logical structure of the consequence argument. To a greater extent, this discussion will be focused on how compatibilists have interpreted "ability" in a hypothetical manner and why I find this interpretation problematic. A central part of this discussion will focus on the work of David Lewis (Lewis, 1981). I will then proceed to show how Frankfurt's account is deeply problematic and why compatibilists ultimately fail to answer the consequence argument sufficiently. By doing so, I attempt to show why I do not find compatibilism worth endorsing and why defenders of compatibilism commonly seem to miss what the free will debate fundamentally is about. My motivation for critiquing compatibilism and showing why I find the view implausible first is to provide a fundament for discussing incompatibilist viewpoints in chapters 2, 3 and 4.

The subject of chapter 2 is agent-causal libertarianism. Particular to this form of libertarianism is that agents can freely cause actions and that the contents of the action and the fact that the agent causes some event are not determined by prior events (Clarke, Capes, Swenson, 2021, 3). I will first provide a more thorough explanation of what this view entails. Following this, I will outline some of the view's strengths and explain how agent-causal libertarianism can account for some objections and problems the view faces. Lastly, I will argue why I do not consider the view plausible enough to endorse. I will discuss several notable agent-causal libertarians throughout this chapter, such as Thomas Reid and Roderick Chisholm. I will give the most attention to the works of contemporary defender, Timothy O'Connor (O'Connor, 1996, 2000). My reasoning for this is because his account builds on the ideas of Reid and Chisholm, and a contemporary account is needed to discuss some of the issues agent-causal libertarianism faces due to modern scientific discoveries.

---

[5] The consequence argument argues that incompatibilism holds true by showing how we do not have free will if determinism were to hold true. Simplified, the argument can be laid out the following way: No one has power over the facts of the past and the laws of nature and no one has power over the fact that the facts of the past and the laws of nature entail every fact of the future. Therefore, no one has power over the facts of the future (McKenna, Coates, 2024, 3.1).

Chapter 3 concerns event-causal libertarianism. This view shares traits with agent-causal libertarianism in that both are incompatibilist theories that claim we have free will. The view also shares traits with compatibilism in that both claim we have free will and as most compatibilist views are also event-causal:

> Compatibilist accounts of free action are typically event-causal views, invoking event-causal accounts of action. The simplest event-causal incompatibilist theory takes the requirements of a good compatibilist account and adds that certain agent-involving events that cause the action must nondeterministically cause it. (Clarke, Capes, Swenson, 2021, 2)

In this chapter, I will focus on Robert Kane's account of event-causal libertarianism (Kane, 2002, 2007). I will focus on Kane's view as he provides an in-depth account, which he attempts to partly ground in empirical science. This is a notable strength, as chapter 2 will illuminate how agent-causal libertarianism is lacking in this regard. I will begin by explaining Kane's account and show how he attempts to answer common questions in the free will debate. Then, I will present a multitude of objections to his view to show why it, despite its strengths, is holistically implausible and why I do not consider his account worth endorsing.

In light of the three previous chapters, chapter 4 begins with a discussion on what type of account on the question of whether we have free will and moral responsibility is worth adopting. I will argue that hard incompatibilism is the most plausible alternative by drawing on Derk Pereboom's work's (Pereboom, 2001, 2007, 2014). Then, I will argue that adopting this view is not a negative outcome and that doing so does not have as detrimental consequences as some have claimed. I will show why I believe this to be the case by showing how many aspects of our sense of morality, personal relationships, and attitudes are not threatened by a hard incompatibilist view. Lastly, I will explore what the implications of lacking moral responsibility would have for our sense of blame and for our justice systems. I do so in order to argue that we can abolish our sense of blame and revise certain aspects of our justice system for the better on a hard incompatibilist view.

1. Compatibilism

1.1 Compatibilism & Frankfurt Cases

Compatibilism is the thesis that free will and determinism are compatible (McKenna, Coates, 2024, foreword). In this chapter, I will examine whether compatibilism can provide a plausible account of free will and moral responsibility. I will proceed by outlining why one might endorse this view, showcase what I consider to be some of the most important arguments in favor of the position, and provide counterarguments to argue why I ultimately do not find this view plausible. In this section I will lay out the key strengths of compatibilism, provide a brief historical context of the view, and present a Frankfurt case, which is a famous argument in favor of compatibilism.

Due to how entwined the concepts of free will and moral responsibility are, most compatibilist accounts imply or directly argue that determinism and moral responsibility are also compatible. The strength of this viewpoint is that if determinism turned out to be true, we could still retain our free will and moral responsibility. In turn, compatibilists avoid the issue libertarian accounts face of explaining how free will can originate in agents or events in a nondeterministic way. Compatibilism can therefore be an appealing position, and the majority of academic philosophers endorse a compatibilist view regarding the question of whether we have free will (Bourget & Chalmers, 2023, p7).[6]

Compatibilism has taken on various forms throughout the years. Classical compatibilism, pioneered by thinkers such as Thomas Hobbes in the 1600s, mainly tried to reconcile free will with determinism by claiming that doing what one wants to do without being hindered constitutes an act that is performed with free will.

> For the classical compatibilist, then, free will is an ability to do what one wants. It is
> therefore plausible to conclude that the truth of determinism does not entail that agents

---

[6] While Bourget & Chalmers surveyed academic philosophers, they did not survey free will specialists. This does raise the question of whether compatibilism would hold a majority vote among free will specialists. Notably, Peter van Inwagen assumed in 2008 that most philosophers are most likely compatibilists and followed this up with the claim that most philosophers with specialized knowledge of free will are, in fact, incompatibilists (van Inwagen, 2008, p337-338).

lack free will since it does not entail that agents never do what they wish to do, nor that agents are necessarily encumbered in acting. (McKenna & Coates, 2024, 2.1)

However, in the contemporary free will discourse compatibilism has become a more nuanced position than classical compatibilism was. This is mainly due to the influence of 3 key works.

Three major contributions in the 1960s profoundly altered the face of compatibilism: the incompatibilists' Consequence Argument, Frankfurt's attack on the Principle of Alternative Possibilities (PAP), and Strawson's focus upon the reactive attitudes. Every resultant compatibilist account in the contemporary literature is shaped in some way by at least one of these influences. (MeKenna & Coates, 2024, 4)

As the abovementioned works are the most relevant and influential for where compatibilism stands today, I will focus my attention on these. In this chapter I will mainly focus on Frankfurt and the consequence argument. Strawson's account of reactive attitudes will be discussed in section 3.4.

Compatibilists provide various answers to how we can have free will and be held morally responsible if determinism were to hold true. Oftentimes, this consists of showing that free will is not the type of concept that requires humans to cause their actions in a libertarian sense but rather that something else is at the heart of what enables us to have free will. As Kane puts it, compatibilist arguments have us reflect on what we ordinarily mean when we say our actions are free (Kane, 2005, p13). Frankfurt cases are one such argument, and I will proceed by outlining it. To discuss Frankfurt cases, I first need to introduce a key concept in the free will discourse: the principle of alternate possibilities. This principle states that in order to be morally responsible for our actions, we need to have been able to do otherwise (Robb, 2020, foreword). Prima facie, the principle of alternate possibilities, henceforth PAP, seems plausible. If a person did not have alternate possibilities, henceforth AP, how can they be held morally responsible for the sole action they had to take? As the PAP makes such an intuitive and robust point about moral responsibility and free will, this principle has been a central point of discussion in the free will debate (Blumenfield 1971, Naylor 1984, Lamb 1993, Copp 1997, Glatz 2008, Alvarez 2009). The PAP also poses a threat against compatibilism because if determinism were to hold true, we would not have alternate possibilities, and we would seemingly not be morally responsible if both the PAP and determinism were to hold true.

Compatibilists have therefore argued that the PAP does not hold true. One of the most influential works in this discussion is Harry Frankfurt's "Alternate Possibilities and Moral Responsibility" (Frankfurt, 1969).[7] In this work, Frankfurt relies heavily on thought-experiments when arguing the point that we can be held morally responsible even when we lack AP. These thought-experiments have since been dubbed Frankfurt cases, which can be laid out in a multitude of ways, such as the following:

> 1. Alicia and Beyonce are roommates and have agreed to take turns vacuuming every other week.

> 2. Alicia has a suspicion that Beyonce is not going to clean the carpet next week, when it is her turn to do so.

> 3. Therefore, Alicia plants a microchip in Beyonce's brain which monitors whether she chooses to vacuum or not.

> 4. If Beyonce does choose to vacuum the chip will not do anything, but if she decides against vacuuming the microchip will control Beyonce to vacuum against her will.

> 5. When Beyonce ends up deciding to vacuum of her own will, it appears as if she bears moral responsibility for doing so, even if she does not have other options available to her. (Mandik, 2013, p177, my wording and formatting)

Due to the microchip, Beyonce will end up vacuuming the floor regardless of what she wills to do. However, despite the result being that she ends up vacuuming the floor no matter what, whether she intends to vacuum of her own volition should seemingly play some role in how we assign her responsibility for the action. Put another way, it would seem unjustified to morally praise or blame Beyonce on the same grounds when she attempts to do a good act and succeeds and when she does so because she was compelled and succeeds.

If we accept that we can morally praise Beyonce if she vacuums the floor of her own volition, despite not being able to do anything but vacuuming the floor, it appears as though we have circumvented the challenge posed by the PAP. This is because she could not have done

---

[7] The PAP has also been an important point of discussion in works that expand upon the works of Frankfurt (Ginet, 1996, Fischer, 2010, Harrison, 2011).

otherwise than vacuum the floor, but despite the lack of alternate possibilities, she could still be held morally responsible. Therefore, we do not need to appeal to libertarian arguments for how we can be held morally responsible, and more importantly, moral responsibility seems to be possible in some cases, even if determinism were to hold true.

1.2 Compatibilism & The Consequence argument

In addition to the discourse surrounding "the ability to do otherwise", another central theme for compatibilism has been to examine what "ability" means more broadly. In particular, what we mean by "ability" has been a central point of discussion surrounding the consequence argument. In this section I will discuss this argument. Given that the argument attempts to show that incompatibilism holds true, I will also discuss how compatibilists have attempted to provide counterarguments against it.

A simplified version of the consequence argument can be laid out as follows:

1. No one has power over the facts of the past and the laws of nature.
2. No one has power over the fact that the facts of the past and the laws of nature entail every fact of the future (i.e., determinism is true).
3. Therefore, no one has power over the facts of the future (McKenna, Coates, 2024, 3.1)

Carl Ginet was the first philosopher to develop the consequence argument in 1966 (McKenna, Coates, 2024, 3). However, various philosophers have developed the argument further. Peter van Inwagen's work on it (van Inwagen, 1975, 1983) has been among the most influential, and his work remains a central point of discussion surrounding the argument (Kapitan, 2001, p128-135, Vihvelin, 2022, 5). The argument is meant to showcase the difficulty of how we can be said to have free will if determinism holds true. This is because the conclusion seems to follow from the premises (which stipulate determinism), and it is difficult to say how we can be said to have free will, if the conclusion that we have no power over the facts of the future holds true. However, showing where the argument goes wrong could, in turn, be a major reason to endorse compatibilism because the argument serves as a fundamental reason why I, van Inwagen and other incompatibilists endorse our position.

Some compatibilists (Lewis 1981, Graham 2008, Perry 2010, Dorr 2016) have argued against how incompatibilists have interpreted the first premise of the consequence argument. This difference in interpretation concerns the word "can", as debates concerning what we "can" and "cannot" do also concern what "having the power or ability to do something" means (Kane, 2005, p26). With an incompatibilist reading of the first premise of the consequence argument, we can interpret the lack of power over the past and natural laws to mean that we would need to have some ability to perform some action that would cause the laws or past to be different if the premise was to be false. One way to read the first premise of the consequence argument on this interpretation is the following.

1. No one has the ability to perform some action that would cause the facts of the past or the laws of nature to be different.

As for how some compatibilists interpret the argument, we need first to consider how "having the power to do something" can be taken to mean "If you wanted or tried to do so something, you would do it" (Kane, 2007, p11-12). This interpretation of ability is called "hypothetical" as it contains "if" in it (Kane, 2002, p27). It is this interpretation that some compatibilists have taken the word "ability" to mean. An example of this type of ability would be for me to possess the ability to sip my coffee, even though I did not will to sip it and therefore abstained from doing so. However, I would have succeeded in doing so if (hypothetically) I willed to sip my coffee and attempted to do so. On this interpretation, I did, therefore, possess an ability to sip my coffee despite not actually sipping it. This is different from the incompatibilist interpretation of ability, where I would be causing the drinking of the coffee if I were able to sip it and did so. This compatibilist interpretation of ability has been referred to as broad ability, whereas the incompatibilist interpretation has been referred to as casual ability.

> Broad Ability: S is broadly able at t to bring about p iff there is a course of action K such that at t (1) S is able to do K, and (2) were S to do K then p.

> Causal Ability: S is causally able at t to bring about p iff there is a course of action K such that at t (1) S is able to do K, and (2) S's doing K would make it the case that p. (Kapitan, 2001 p134)

With this distinction between ability claims in mind, I will consider the two main ways compatibilists have argued against the consequence argument.

The first way has been to show that the "Transfer of powerlessness Principle", also known as Rule Beta, does not serve the type of role in the consequence argument that incompatibilists claim it does. The Transfer of powerlessness Principle can be laid out as follows.

> (TP) If there is nothing anyone can do to change X, and if Y is a necessary consequence of X (if it must be that, if X occurs, Y occurs), then there is nothing anyone can do to change Y. (Kane, 2007, p11)

In the case of the consequence argument, this principle shows how our powerlessness over the facts of the past and the laws of nature transfers to our powerlessness over the facts of the future.

> 1. If there is nothing anyone can do to change the laws of nature, the past, or the fact that they entail every fact of the future. And no one having power over the facts of the future is a necessary consequence of this.

> 2. Then there is nothing anyone can do to change the fact that no one has power over the facts of the future.

However, with a broad interpretation of ability, our powerlessness over the past and laws does not necessarily carry over from the premises to the conclusion of the consequence argument, rendering the inference invalid.

To illustrate why, consider how an unconstrained and coherent person in an everyday setting could have chosen differently. Kane uses the example of Molly, who raises her hand but could have acted differently because "she would have done otherwise than raise her hand if she had chosen or wanted to do otherwise" (Kane, 2005, p27). With a hypothetical analysis of ability, the powerlessness over the past and the laws will, in this case, not have been transferred to a powerlessness over the future. This is because despite not having power over the laws or the past, Molly still holds power over the future in that she could have abstained from raising her hand, if she had willed to do so. Therefore, she hypothetically could have changed the future by keeping her hand lowered, and she holds power over the facts of the future for this reason.

This point is illustrated further by Tomas Kapitan, who points out how the consequence argument rules out our ability to do some things we usually are assumed to have the ability to do (Kapitan, 2002, p129). Under ordinary circumstances, raising or not raising one's hand is something we have power over in a different sense than changing the laws of nature or the past. I seem to have the ability to take a sip of my coffee or abstain from doing so, but I do not seem to have the ability to change the past or the laws of nature. This seems intuitive, but this difference in ability does not refute premise 1 or 2 of the consequence argument. However, despite the premises of the argument remaining true on this reading (The past and the natural laws are fixed, they entail every fact of the future, and we cannot change that), the conclusion nevertheless comes out false (Kane, 2005, p28). This is because even if determinism holds true, Molly "can" still raise her hand or lower it, and I "can" still sip or abstain from sipping my coffee. For this reason, in one sense, we have power over the facts of the future, as exercising the abovementioned ability would change it. By contrasting this ability with things we cannot do, such as changing the laws or the past, we can see how the consequence argument seems to claim that changing the laws or past is something we are just as incapable of as keeping a hand lowered or sipping a coffee.

Compatibilists have also argued against the transfer of powerlessness principle for a different reason. Namely, they have argued that Rule Beta is invalid to begin with. This has proven successful, as rule Beta and the way it is utilized in the consequence argument has widely been accepted as false as it entails agglomeration (Vihvelin, 2022, 5). Agglomeration is a logical principle, originally laid out by Bernard Williams, which applies Kant's principle that "ought implies can" to two distinct options one ought to do, inferring that one ought to do both of these options (Williams, 1965, p118). The agglomeration principle is controversial, as we can imagine cases where the two distinct options are mutually exclusive. For example, we can imagine a situation where one ought to go to two different places at the same time or vote for two separate parties. This raises the question of how one ought to do both things when it is impossible to do both simultaneously. In turn, it can be questioned whether this type of inference can be said to be valid. In van Inwagen's consequence argument, agglomeration becomes problematic due to the claim that "no one has, or ever had, any choice about" two distinct premises p and q. This is problematic as it opens up the possibility of using mutually exclusive premises in the place of p and q and retaining that no one has, or ever had, any choice about whether p nor q. Using a more

mundane example of tossing a coin makes this issue more apparent, as the coin toss has two contrasting outcomes.

> let 'p' abbreviate 'The coin does not land heads', let 'q' abbreviate 'The coin does not land tails', and suppose that it's a fair coin which isn't tossed but someone could have tossed it (Vihvelin, 2005, 5).[8]

The result of using Rule Beta in this manner is that it follows that p and q, and no one has, or ever had, any choice about whether p and q. In the case of the coin, it might land either heads or tails, and we might not have a choice in which event will occur, but it cannot be said that both events will occur. Applying the coin example in the format of the consequence argument, we can see how agglomeration is not valid due to how the second premise then makes the claim that two mutually exclusive outcomes would occur.

1. No one has the ability to perform some action that would cause some coin to land heads
2. No one has the ability to perform some action that would cause some coin landing heads to not entail some coin landing tails
3. Therefore, no one has power over some coin landing tails

One reason one might endorse compatibilism is the invalidity of Rule Beta and its bearing on the consequence argument. This leaves incompatibilists with the choice of finding a different way of showing how a lack of power over the laws and past entails a lack of power over the future or providing a different way to argue for their position altogether.

The second strategy that compatibilists have employed against the consequence argument is to disprove the first premise and show that we do, in fact, have power over the past or the laws of nature. I will proceed by illuminating how compatibilists have argued this point. Falsifying the first premise on a casual interpretation of ability would require showing how we can be said to have an ability to cause the past or laws of nature to be different with some action, which seems impossible. However, with a broad interpretation of ability, some compatibilists have attempted to refute the fact that we lack power over the past. In order to make this claim, compatibilists have distinguished between the ability to change the past and the ability to act in a manner where

---

[8] Thomas McKay and David Johnson originally laid out this example of tossing a coin (McKay, Johnson, 1996), and this is Vihvelin's wording of their example.

if one were so to act, the past would have been different (McKenna, Coates, 2024, 4.1.1). Similar to the argument laid out above about how different interpretations of "can", "power", and "ability" can render the conclusion of the consequence false, this argument relies on a similar interpretation to refute the first premise. The line of reasoning behind this argument can be laid out as follows:

> Consider the claim, *If I were dancing on the French Riviera right now, I'd be a lot richer than I am*. Certainly this claim does not mean that if I go to the French Riviera to dance, I will *thereby* be made richer. It only means that were I to have gone there to tango, I would have to have had a lot more cash beforehand in order to finance my escapades. (McKenna, Coates, 2024, 4.1.1)

In other words, by interpreting the "power over the facts of the past" on a broad analysis of ability, the above quote shows how applying a similar type of logic to a more mundane example results in a plausible statement rather than the claim that we could cause the past to be different. Similarly, maybe I was determined to sip my coffee, but if the past was different, I would have had the ability to abstain from sipping it if I willed to do so. On this interpretation, the first premise of the consequence argument can be considered refuted. This is because we do, in a sense, have power over the past.

A similar compatibilist counterargument against the first premise of the consequence argument has been made by refuting that we do not have power over the laws of nature. To make this claim, compatibilists have drawn a distinction between the ability to perform acts that violate laws of nature and the ability to perform acts so that if they were to be carried out, a law of nature that normally obtains would not obtain (McKenna, Coates, 2024, 4.1.2). David Lewis, a notable defender of this interpretation, has labelled these two ability notions as strong and weak, which resembles the distinction between casual and broad abilities discussed above.

> (Weak Thesis) I am able to do something such that, if I did it, a law would be broken.

> (Strong Thesis) I am able to break a law (Lewis, 1981, p115)

This line of reasoning distinguishes between a strong ability of acting in a way that violates laws of nature by causing a law to be broken, and a weak ability where the laws of nature that obtain would not have obtained if some action was performed. Lewis points out how talking of ability

on a weak interpretation, such as in the case of a simple act like raising or keeping a hand lowered, is not an instance of inability, even if the act was determined (Lewis, 1981, p113). In other words, even if I was determined to keep my hand lowered, I would have the ability to raise my hand had I willed to do so. A law would have needed to be broken, but if a law had been broken, I could have raised my hand without me having to cause that law to be broken. This way, our ability to do something can be understood in a way where the first premise of the consequence can be disputed. We need not actually break any laws, or even be able to break them. On a weak interpretation of ability concerning laws, we can instead be said to have an ability to perform actions such that laws would not have obtained if we did perform some such action.

> Thus I insist that I was able to raise my hand, and I acknowledge that a law would have been broken had I done so, but I deny that I am therefore able to break a law. To uphold my instance of soft determinism, I need not claim any incredible powers. To uphold the compatibilism that I actually believe, I need not claim that such powers are even possible (Lewis, 1981, p117).[9]

Therefore, even if determinism holds true, we can, according to Lewis, be said to have some type of power over the laws of nature without having to make the outlandish claim that we can break them.

The abovementioned ways of interpreting "ability" in a broader and weaker sense can be used to advocate for the compatibility of determinism and free will. Some accept the broader interpretations while others do not, and as a result, this disagreement in interpretation tends to lead to an impasse in the discourse surrounding the consequence argument (Kane, 2005, p30, 2007, p13). Despite the difficulty of circumventing this standstill, the consequence argument remains relevant in the debate between incompatibilists and compatibilists. I have examined several compatibilist counterarguments against it, but for the incompatibilist, we will come to see how these arguments can be regarded as insufficient for several reasons.

---

[9] Soft determinists are compatibilists who believe determinism does not undermine the forms of free will worth having (Kane, 2005, p69). Compatibilists do not have to be soft determinists because they can be antagonistic about whether determinism holds true (Vihvelin, 2022, 1). This results in compatibilists being prone to some arguments against soft determinism but not others (Lewis, 1981, p113), which is why Lewis draws this distinction.

1.3 The issues with Frankfurt Cases

Previously, I highlighted how Frankfurt cases attempt to show how we can be held morally responsible without alternate possibilities. However, I remain skeptical about whether we can be said to be morally responsible in virtue of Frankfurt cases, and I will proceed by explaining why I hold this view. Due to Frankfurt cases, many compatibilists have drawn the moral that moral responsibility and the kind of free will required for it are compatible with determinism (Mandik, 2013, p178). This highlights a central issue with compatibilism in general but one that I find particularly detrimental when it comes to Frankfurt cases. Namely, if one is to remain agnostic about whether determinism holds true, while maintaining that if determinism turned out to hold true it would be compatible with free will, any argument in favor of this view must still be valid if we stipulate determinism. Therefore, if one is to claim that Frankfurt cases show that we can be held morally responsible regardless of whether determinism holds true or not, Frankfurt cases must also show that we can be morally responsible if determinism holds true.

Personally, I find it hard to see how Frankfurt cases would serve as an argument for moral responsibility if determinism turned out to hold true. Vargas shares this sentiment with me, as he writes that he finds that it is not obvious that we should view the case as one where there is moral responsibility if we stipulate determinism and that the original cases rely on being under-described (Vargas, 2007, p134). Let us say that determinism holds true, and in the above-mentioned Frankfurt case Beyonce ends up vacuuming the floor without interference from the microchip. The question, then, is how this action is any different from her performing any everyday action, such as raising her hand or sipping her coffee? She was not coerced nor manipulated, and she "chose" to vacuum the floor herself. However, there seems to be nothing about the Frankfurt case that separates the vacuuming of the floor from any everyday action that runs into the same issue of how we can be held morally responsible in a deterministic world.

In some deterministic world, Alica was determined to plant the microchip in Beyonce, and Beyonce was determined to vacuum the floor without interference from the chip. In another world where determinism also holds true that closely resembles the abovementioned one, Beyonce would have vacuumed the floor due to the microchip interfering and making her vacuum the floor despite her attempting to do something else. In this second deterministic world,

Alica was determined to plant the microchip in Beyonce, and Beyonce was determined to vacuum the floor with interference from the chip. Suppose we morally praise Beyonce for vacuuming without inference from the chip in one world and blame her for merely doing so due to interference of the chip in another. In that case, it also needs to be pointed out what separates these actions from any ordinary day action, which would also be determined. In other words, if Beyonce "willed" to vacuum of her own volition and deserves moral praise for this, how is action different from any action she performs that she is determined to will? Due to the difficulty of answering this question, I agree with Ted Wakefield that compatibilists face some non-trivial difficulties in this discussion, as pointing to alternate possibilities in a Frankfurt case also involves giving up on the claim that there would be responsibility without alternate possibilities (Wakefield, 2003, p621).

1.4 Defending the Consequence Argument

Discussions like these about Frankfurt cases and the "ability to do otherwise" lead us back to the discussion on how to interpret "ability", "can" and "power" to begin with. As previously shown, a broad compatibilist interpretation of "ability" could account for how we can still have the ability to do something else than what we are determined to do if determinism holds true. However, to achieve this, I also believe the consequence argument would need to be successfully circumvented, as the argument shows why we could not have done otherwise if the ability to do so is to be equated with having power of the future. In my view, the compatibilist arguments against the consequence argument suffer from several issues that I will now address. Compatibilists have argued against the consequence argument in two main ways. The first way was to argue that the transfer of powerlessness principle renders the inference in the argument invalid, and the second way was to argue that the argument takes on a different meaning with a hypothetical analysis of "can". I will proceed by addressing the first way.

As previously mentioned, Rule Beta has been accepted as uncontroversially invalid. Van Inwagen has also conceded that Rule Beta is invalid (van Inwagen, 2002, p161). Furthermore, he has also conceded that the version of the consequence argument that he laid out, which utilizes

Rule Beta, is invalid (ibid, p165).[10] However, van Inwagen argues that this issue can be circumvented by reformulating how powerlessness is expressed in versions of the consequence argument that utilize Rule Beta. Initially, powerlessness was expressed the following way in these versions of the argument: "No one has, or ever had, any choice about whether something holds true" (van Inwagen, 1983, p93, p98, p149). However, to avoid an invalid inference, this phrase can be reformulated to "if p, no one has the power to do anything that, if they did it, p would be false". Kadri Vihvelin expands on how van Inwagen argues this point the following way, where Np denotes "p, and no one has, or ever had, any choice about whether p":

> replacing 'N' with 'N*', where 'N*p⋆' says "p and no one can, or ever could, do anything such that if she did it, p might be false". Agglomeration is valid for 'N*', and thus this particular objection to the validity of Beta does not apply. (Vihvelin, 2022, 5)

If we again use the tossing of a coin as an example, we can see how N* does not entail agglomeration in the same way N does, and we can lay out the coin example in the format of the consequence argument as follows:

1. No one can, or ever could, do anything such that if she did it, some coin landing heads would be false.
2. No one can, or ever could, do anything such that if she did it, some coin landing heads would not entail some coin landing tails.
3. Therefore, no one can, or ever could, do anything such that if she did it, some coin landing tails might be false.

On the surface, this use of N* can seem similar to the use of N, as the second premise still sounds problematic. However, instead of the second premise being contradictory due to N, N* yields a premise that says something about our inability to cause a contradictory result. A coin landing heads also entailing it landing tails is unquestionably impossible. N says that there is nothing we can do to avoid this contradictory result. N* says there is nothing we can do to cause this contradictory result. This way, N* avoids the issue that agglomeration initially posed for the

---

[10] Specifically, van Inwagen writes that McKay and Johnson pointed out how rule Alpha and Rule Beta together imply agglomeration (van Inwagen, 2002, p160).

consequence argument. Using N* instead of N, the consequence argument can then be formulated as follows:

1. No one can, or ever could, do anything such that if she did it, the facts of the past or the laws of nature would be different.
2. No one can, or ever could, do anything such that if she did it, this action would cause the facts of the past and the laws of nature to not entail every fact of the future
3. Therefore, no one can, or ever could, do anything such that if she did it, the facts of the future would be different

In my opinion, this formulation preserves the original sentiment of the argument, as it plausibly shows how, given determinism, we would not have power over the future. Therefore, I agree with van Inwagen when he claims that using N* instead of N sufficiently answers compatibilists who have argued against the validity of the consequence argument.

> What McKay and Johnson's counterexample shows is that the concept "not having a choice about" has the wrong logical properties to capture the idea I wanted to capture-the idea of the sheer inescapability of a state of affairs. But if "N" is redefined in the way I have proposed, the redefined "N" does capture this idea. (van Inwagen, 2002, p166)

As this new version of the argument does not utilize Rule Beta, but still retains the sentiment of the original argument while remaining logically sound, I find it to be a sufficient counterargument to the compatibilist claim that the consequence argument is invalid due to it utilizing Rule Beta and therefore entailing agglomeration.

The second way compatibilists have attempted to refute the consequence argument is by interpreting the notion of ability as hypothetical. Specifically, compatibilists have attempted to refute the first premise that we do not have the power over the laws of nature, claiming that on a hypothetical analysis of "can", we possess abilities that can be said to refute this premise. My main issue with this line of reasoning is that this interpretation misses what the first premise attempts to state. This sentiment is similar to what van Inwagen expresses in the quote above. I claim that the "The sheer inescapability of a state of affairs" that van Inwagen is referring to is not shaken to any degree by a hypothetical analysis of ability. To illustrate my point, let us first examine how the initial plausibility of the consequence argument and its first premise is discussed in the free will discourse. Kane writes that the argument seems initially plausible

(Kane, 2007, p11) and Kapitan, who argues in favor of a compatibilist interpretation of ability, concedes that, at the face of it, refuting the first premise seems like an absurd approach (Kapitan, 2002, p132). I also agree with Manuel Vargas that the naturalness of reading the consequence argument in an incompatibilist manner captures an important part of how we naturally think about the issues surrounding ability:

> Several critics have focused on how an antecedently compatibilist reading of the relevant ability term (for example, the ability to break the laws of nature) makes the argument unpersuasive against antecedently committed compatibilists. Suppose these critics are right. If so, it would mean that Consequence-style arguments couldn't rule out the possibility of compatibilism. From this, we might think things look like a standoff, at best, and at worst a real problem for incompatibilism. Here, the innocuous issue has some role to play: the "naturalness" or ease of the incompatibilist reading of the argument is itself evidence that the argument captures an important part of the contents and logic of commonsense thinking about these issues. Even if the contents of the argument cannot rule out a compatibilist reading, the naturalness of the incompatibilist readings of the argument strongly suggests that we can and do understand these issues in incompatibilist ways, at least sometimes. (Vargas, 2007, p132-133)

The above statements illuminate why I believe compatibilists have needed to go to great lengths to even begin refuting the consequence argument and have in the process moved past what the argument intuitively expresses. I do not refute the logic of how the consequence argument would be invalid on a hypothetical analysis of can. Instead, I refute that this notion of ability applies to what the argument is attempting to state.

This idea is captured more aggressively in Kane's formulation of how defenders of the consequence, such as van Inwagen, might take issue with this hypothetical analysis and how they might respond to such a claim. I do not endorse libertarianism in the same way as Kane. However, I fully agree with his overall sentiment on this topic as it captures the essence of why I endorse incompatibilism instead of compatibilism:

> ..your compatibilist analysis was rigged in the first place to make freedom compatible with determinism. On your analysis, persons can jump the fence even though their doing so here and now is impossible, given the past and the laws of nature. That is not what we

libertarians mean by 'can' in the Consequence Argument. We mean it is possible that you do it here and now, given all the facts that presently obtain. If your analysis allows you to say that persons can do otherwise, even though they can't change the past and the laws of nature and even though their actions are a necessary consequence of the past and the laws of nature, then something must be wrong with your compatibilist analysis." (Kane, 2007, p13)

I agree something must be wrong with this compatibilist analysis. What is wrong is the lack of regard for the inescapability of the state of affairs; the "here and now" where hypotheticals are not realized.

I believe any analysis that relies on a hypothetical ability notion has a fundamental flaw if it does not properly account for this inescapability. Therefore, I also believe this flaw is detrimental to arguments that have attempted to refute that we lack power over the past and arguments that have attempted to refute that we lack power over the natural laws. While there are a multitude of works worth addressing on this topic (Dorr 2016, Perry, 2003, 2010) this is also a topic where any one argument requires a thorough analysis. I have therefore focused my attention on Lewis' position that we have power over the laws. However, by critiquing his work I am also looking to show why other arguments that rely on a hypothetical analysis of can are problematic for similar reasons.

As mentioned earlier, Lewis distinguishes between a strong ability to break a law and a weak ability to act in a way where a law would have been broken if they had acted in such a way. I sympathize with this distinction, as on this weaker, broader interpretation of ability, we can speak of how actions relate to the laws in an intuitive manner. For example, I could take a sip of my coffee or choose not to. I ended up not doing so, but it feels like I would have been able to take a sip if I decided to do so. Nothing would seemingly have stopped me if a law was broken, so that I ended up willing to take a sip instead. In so far as we are talking about what we could have done if a law is broken, this ability notion is unproblematic. However, Lewis argues that this would be the case even if determinism held true and that I would not have been inable to take a sip of my coffee even if I was determined not to do so (Lewis, 1981, p115-116). This is where I take issue with Lewis' view. I say this because if I were determined to perform some specific action, it would not have happened that I performed some other action. I concede that if

a law was broken and this led to me performing some other action, I would have performed that action. But something causing a law of nature to be broken will not happen. If a law is not determined to hold, I see no reason why it should be called a law of nature, and if a law is determined to hold, it will not be broken.

A law being broken prior to the action would require a local, divergence miracle (Lewis, 1981, p117).[11] Lewis' claim is more so that we can perform actions so that a law that obtains would not have obtained, rather than claiming that local miracles will occur so that this action will happen. However, if determinism holds true and the laws and past led to me abstaining from sipping my coffee, the fact remains that it is impossible that I would have sipped my coffee. I might have sipped my coffee if a law had been broken beforehand, but a law would not have been broken beforehand. Compatibilists and incompatibilists will disagree on what being able to sip my coffee means here, but on both analysis of "can", what will happen nevertheless seems inevitable. For incompatibilists and compatibilists, the meaning of "Could have done" is different, but if determinism holds true, the "Will end up doing" is the same.

This creates a problem in Lewis' distinction between a "commonplace ability" to break a window, something he says he is able to do if he threw a stone in the direction of a window, and a "marvellous ability" to break a law, an ability Lewis says he cannot credibly claim. (Lewis, 1981, p115). However, the following argument showcases why I find this distinction problematic:

> 1. If A is not determined to happen and determinism holds true, it is impossible that A will happen.
>
> 2. Nothing is equally impossible.
>
> 3. Therefore, if determinism holds true, all actions and events that will not happen are equally impossible.

---

[11] Lewis' viewpoint is also known as local miracle compatibilism (Oakley, 2006, p337). The proponent of local miracle compatibilism argues that sometimes agents are able to do something such that an actual law of nature would not have been a law and that local miracle counterfactuals are sometimes consistent with ability-claims (Pendergraft, 2011, p251).

In other words, Lewis breaking the window and him breaking the laws of nature are equally impossible if determinism holds true. The latter would require a local miracle as opposed to the first, and both would require a different world, but it is nevertheless impossible for either outcome to occur in this world. Presumably, breaking a window would only require a slightly different world, whereas the breaking of a law would require a vastly different world where a local miracle occurs. However, if an action or event occurring requires a different world, what difference does it make how different that world is from ours? What difference does it make if it requires a local miracle? I claim none, as Lewis' distinction between marvellous and commonplace abilities makes no difference in a world where acting on these "abilities" is impossible. Lewis claims he is not inable to break the window in virtue of him having succeeded in doing so if he willed to do so. I claim he is inable in virtue of determinism being stipulated and that it, therefore, is impossible for him to will to do so. This is my issue with the hypothetical analysis of ability. It is an interpretation that claims that we can do what is impossible when determinism is stipulated. I do not think we can reasonably be said to be able to do what is impossible. This is why I have a fundamental issue with any argument that relies on a hypothetical interpretation of ability to argue in favour of compatibilism.

I will proceed by critiquing Lewis' definition of ability from a different angle. To make my point, we first need to consider how Lewis views the relation between certain ability claims and counterfactuals when he critiques an incompatibilist reading of the consequence argument:

The argument trades on an equivocation between two counterfactuals.

(C1) If I had raised my hand, the laws (or the past) would have been different.

(C2) If I had raised my hand, I would thereby have *caused* the laws (or the past) to be different.

There is a corresponding equivocation between two ability claims:

(A1) I had the ability to do something (raise my hand) such that if I had exercised my ability, the laws (or the past) would have been different.

(A2) I had the ability to do something (raise my hand) such that if I had exercised my ability, I would thereby have *caused* the laws (or the past) to be different.

The problem with the argument, says Lewis, is that it equivocates between these two ability claims. To count as a reductio against the compatibilist, the argument must establish that the compatibilist is committed to (A2). But the compatibilist is committed only to (C1) and thus only to (A1). (Vihvelin, 2022, 5, my formatting)

Here, Lewis' argument concerns the relation between counterfactuals and ability claims. The argument is an attempt to circumvent the claim that we would need to casually break a law to not be under the power of the laws of nature. However, my issue with this argument is related to how he defines the ability claims themselves, not how he claims they relate to different counterfactuals. As previously mentioned, I agree that if I had raised my hand instead of keeping it lowered as I did, the laws or the past would have been different. Therefore, C1 is an unproblematic counterfactual. However, I find that the ability claim of A1 itself is deeply problematic. In order to say anything at all, A1 simultaneously has to make a claim about what one can do in this world and a claim about what one could do if the world was different. In this world, it will not happen that I raise my hand as I was determined not to do so, given that we stipulate determinism. I was led to not do so by the past and laws of nature and no local miracle will occur to break the laws to stop this from happening. A local miracle might occur in a different world, and a law might be broken, and I might raise my hand. But in that world, there must also be a determined unfolding of events where a law is broken due to a local miracle.

This raises some questions that are difficult to answer. If the breaking of a law leads to me raising my hand in that world, could that "law" even be called a "law" if it is breakable? How can it be called a "law" if it can be broken, and if it is not a law how are we to make sense of Lewis' claim that a law was broken? Even if we look past the abovementioned problem for the sake of argument, is me raising my hand in that world not an action that is determined by the past and laws all the same? A law was broken due to a local miracle, but as we have stipulated determinism, is this miracle that causes the law to be broken not inevitable as well? If it was inevitable, how are we any less at the mercy of the past and laws of nature? Some law was broken, but something must still dictate why I raised my hand. Is not the negation of the law that would otherwise have led me to abstain from raising my hand just another way to be under the mercy of the laws?

Due to the difficulty of answering these questions, I find Lewis' account problematic. If we stipulate determinism, the laws and the past must determine whether a hand gets raised or not. Even if there is a local miracle, some other set of laws must come into play, or the breaking of the law must serve as a new law that negates what the broken law originally stated. Either way, having to stipulate a possible world and a local miracle without sufficiently accounting for what breaking a law would bring with it makes Lewis's view deeply problematic. He argues in favor of only being committed to A1 as opposed to A2. However, even if he is granted this, his commitment to A1 is deeply problematic for the abovementioned reasons.

In conclusion, the laws, in combination with the past, dictate what we end up doing, regardless of how we interpret "can". If the laws were different, they would, in combination with the past, lead me to do one specific action. In other words, the laws and the past exert a power over us. It would be a different type of power if the laws were different or a law was broken, but clearly a power nonetheless. The lack of regard for this power is what I believe is wrong with the compatibilist analysis of "can" that Kane expressed on behalf of van Inwagen. I argue that any compatibilist argument that is to be taken seriously needs to account for this problem, as this sentiment is what lies at the heart of the consequence argument.

Because of this, I find that the core issue with compatibilism can oftentimes be traced back to the problem of how the past and laws determine the future if determinism holds true. Frankfurt cases fail to show how we can be held morally responsible if we stipulate determinism. The consequence argument cannot utilize Rule Beta, but it can unproblematically be reformulated to account for how determinism poses the same type of threat it always has done. Lewis and others who rely on a hypothetical analysis of "can" are forced to make the claim that we have the ability to do something impossible. This is the core reason why I do not endorse compatibilism. If determinism holds true, we are constrained from freely willing to do otherwise, and it is impossible for us to will anything else than what we will. As this is the case, I do not believe we can be said to have free will to any meaningful degree if determinism holds true, and the compatibility of free will and determinism is therefore also not attainable.

2. Agent-Causal Libertarianism

2.1 Defining Agent-Causation

Given that I do not find the compatibility of determinism and free will attainable, I will proceed by considering incompatibilist accounts of free will. The subject matter of this chapter will be to examine and critique agent-causal libertarianism. In this first section, I will first outline the concept of agent-causation. I will also provide a brief historical description of the concept and illuminate why the question of whether we have agent-causal powers has major implications for both the free will discourse and everyday life.

Defenders of agent-causal views claim that an agent, meaning a person who performs some action, can also be the cause of the action being performed. As Clarke and colleagues put it, the agent is the originator of his or her actions, and it is the agent as a substance that causes the action, rather than an event causing the action (Clarke, Capes, Swenson, 2021, 3.1). Clarke has also pointed out that if agent-causation holds true, the agent can then be regarded as an uncaused cause of her acting (Clarke, 1993, p191). The view differs from an event-causal libertarian view, where an agent's action is wholly reducible to mental states and other events that are involved in the causation of an action (O'Connor & Franklin, 2022, 2.5). According to an agent-causal view, actions are not reducible to events in this manner. It is not the internal or external event the agent is involved in that causes the action; it is the agent themself that, in one way or another, causes the action.

Agent-causation has almost exclusively been thought of as an incompatibilist idea, and those who argue that the position holds true are mainly libertarians.[12] Agent-causation also plays a crucial role in Derk Pereboom's hard incompatibilist account. I say this as his Pereboom believes that we can only be held morally responsible if we have a type of libertarian agent-causal power (Pereboom, 2007, p101) and that agent-causation is the one indeterministic position that would yield a coherent conception of moral responsibility (Pereboom, 2001, p129). In other words, agent-causation holding true is, for Pereboom, a prerequisite for us having moral responsibility, and having moral responsibility in virtue of having agent-causal powers could be conceptually

---

[12] Agent-causation has also been discussed as an option for compatibilism by Ned Markosian (Markosian, 1999).

possible. In turn, if it turns out that humans do not have agent-causal powers, it also follows that we do not have the type of free will required to be held morally responsible. Therefore, as Pereboom deems it unlikely that we have agent-causal powers (Pereboom, 2007, p85, p111-112), this seeming lack of agent-causal powers is foundational for his skeptical view.

Pereboom expands on this idea of agent-causation pointing out how a lack of determinism and a lack of causation among events would make us have free will if we do in fact have agent-causal powers.

> free will of the sort required for moral responsibility is accounted for by the existence of agents who fundamentally as substances have the power to cause actions without being causally determined to do so (e.g., O'Connor 2000, 2008; Clarke 2003). The causation involved in free action is not reducible to causation among events, and what ensures this is that the agent, fundamentally as a substance, has the power to cause an action. For an action to be free it's crucial that the agent who causes the action is not causally determined to cause it. (Pereboom, 2017, 1.4)

Pereboom's point that an agent as a substance causing an action without being determined to do so is necessary for being held morally responsibility is an idea that has been echoed throughout history. This idea can be traced as far back as Epicurus, who favored an agent-causal view and believed that humans were morally responsible for their actions in virtue of this agent-causal power.

> ..some things happen of necessity, others by chance, others through our own agency. For he sees that necessity destroys responsibility and that chance or fortune is inconstant; whereas our own actions are free, and it is to them that praise and blame naturally attach. (Epicurus, 2014)

The similar idea that if our actions are neither determined nor caused by events outside ourselves, we would have free will in virtue of causing our own actions is another idea with ancient roots. Most notably, Aristotle wrote the following.

> "If it is manifest that a man is the author of his own actions, and if we are unable to trace our conduct back to any other origins than those within ourselves, then actions of which

the origins are within us, themselves depend upon us, and are voluntary." (Aristotle, 1997, 3.5, p62)

With these points in mind, we can start to form a clearer picture of what agent-causal libertarianism entails and why the idea has played such an essential role in the free will debate for more than two millennia.

If we do have agent-causal powers, this power would serve as a robust explanation for how we can hold each other morally responsible as we ourselves would be the first causes of our actions. If we do not have agent-causal powers, it seems difficult to explain how we can be said to have free will as we would seemingly not be causing our own actions. Therefore, in contemporary free will debates, agent-causation remains a critical topic. In addition to Pereboom, who finds agent-causation conceptually possible but unlikely to hold true, other thinkers have outright advocated for the view (Chisholm 1976, O'Connor 2000) while others have critiqued it (Kane 2002, Feldman, Buckareff 2003).

We often assign praise and blame in everyday life in a manner that, at the very least, is compatible with agent-causation holding true. On a stronger interpretation, it might even be said that in everyday language, we assume that we have agent-causal powers when we assign moral responsibility to others. We use expressions like "That person did it!" and "That person is guilty!" when referring to other people as the causes of events when we hold them accountable. In turn, we also base our justice systems around this idea and punish those we deem guilty of performing criminal acts. Therefore, agent-causation is worth examining, as whether we have agent-causal powers or not has significant implications for our sense of free will and moral responsibility, as well as our social conduct and our justice systems.

2.2 Defending Agent-Causation

In this section, I will outline some of the most noteworthy arguments in favor of agent-causal libertarianism. I will provide a general account of why one might endorse this view and what some of the key strengths of the position are. Despite the idea of agent-causation having been discussed in various forms all the way back in ancient Greece, 16th century philosopher Thomas Reid has been credited as the developer of agent-causal theories of free will (Nichols, Yaffe

2014, foreword). Reid believed that happenings could be either agent-caused or event-caused and emphasized that actions caused by humans differed from event-caused happenings. In other words, he argued that human actions differed from mere events and that this difference should serve as grounds for accepting that we have agent-causal powers.

Reid's account of why we ought to embrace an agent-causal view is closely tied to his concept of power. He regarded power as a distinct sort of mental effort which we exert when we will to perform an action, and that to have power we must also be able to exert it (ibid, 4.2). Reid argued that the only beings that can exert power are creatures with minds, which gives humans the ability to cause actions that differ from event-causal happenings in the world. Reid also claims that having power is what gives us "moral liberty" (ibid, 4.2). To Reid, it is our minds having the power to will something and us being able to exert this mental power to successfully perform actions that gives us a unique form of ability. To Reid, having this ability means we have agent-causal powers, which in turn makes it possible for us to be held morally responsible. In arguing this, Reid attempts to show how agent-causality could account for how we can have free will and moral responsibility in an otherwise event-causal world.

A later defender of agent-causation, Roderick Chisholm has built upon Reid's account of how having a mind makes it possible for us to exert agent-causal powers. He also builds on Aristotle's abovementioned idea that we need to cause our actions if we are to act freely. However, Chisholm specifies that for a person to be responsible for an action, at least one of the causes of his actions must be immanent, meaning it was caused by the person (Chrisholm, 1964, p9).[13] In explaining how agent-causation of this kind could occur, Chisholm distinguishes between A causing a brain event to happen and a brain event causing A to do something (ibid, p10). Any action a person does is complex; I drank my coffee because I lifted the cup to my mouth with my hand, because I desired coffee, because this desire occurred in my mind. However, Chisholm argues that agent-causation could occur at some part of a causal chain of this kind and that agent-causation presumably happens in the brain (ibid, p8-p10). In other words, me raising my hand and drinking my coffee might be events, but I also caused my brain to have some state that, in turn, brought these events that caused other events about.

---

[13] For the sake of clarity, I will refer to immanent causation as agent-causation. I do so because there is no meaningful distinction between the two for the present purpose of the discussion.

It is uncontroversial that brain states and occurrences in the brain play a part in our actions. However, Chisholm's account still raises the question of how agent-caused brain states differ from event-caused ones. To account for this, he argues that agent-causation is no less clear than event-causation when it comes to this issue.

> "What is the difference between saying, of two events A and B, that B happened and then A happened, and saying that B's happening was the cause of A's happening?" And the only answer that one can give is this-that in the one case the agent was the cause of A's happening and in the other case event B was the cause of A's happening. The nature of transeunt causation is no more clear than is that of immanent causation (Chisholm, 1964, p10-11).[14]

In other words, even if agent-caused brain states seem elusive or mysterious, it is no less elusive than explaining how we can say that an event caused another event. By making this claim, Chisholm echoes Hume's account of our inability to derive causes from observing the external world. However, building on Hume's and Reid's accounts, Chisholm takes his defense of agent-causal libertarianism one step further. Given that Hume might have shown that we cannot derive causation from our perception of external things, we may instead derive it from our own experiences to produce certain effects. In turn, Chisholm argues that if we did not understand agent-causation we would not understand event-causation either (ibid, p11).

In other words, it is our understanding of our own causal tendencies that can provide us with an account of causation at all. Chisholm, therefore, not only provides an account of how the agent makes a brain event happen which serves as a form of uncaused first cause that can cause further events. He also claims that understanding our agent-causal tendencies enables us to understand event-caused events. If we are limited in how we can understand events that cause events, we can, at the very least, understand our own experiences of producing certain effects and thus not only have an account of agent-causation, but of causation at all.

---

[14] Chisholm is referring to event-causation when using the term transeunt causation.

2.3 Answering Counterarguments Against Agent-Causation

In this section, I will show how modern agent-causal libertarianism has expanded upon the ideas laid out in the previous section and how defenders of the position answer counterarguments against their views. The most noteworthy contemporary defender of agent-causality, Timothy O'Connor, emphasizes willingness and intention to perform actions as the ground for acting in ways that can be regarded as agent-caused, which is an idea that was also advocated for by Reid.

> We should first note with Thomas Reid that a particular that freely and directly brings about an effect has to be an agent that can represent possible courses of action to himself and have desires and beliefs concerning those alternatives. Agent causes bring about immediately executive states of intention to act in various ways. (O'Connor, 2000, p72)

In building on Reid's theory and emphasizing the role of intentions, O'Connor accounts for a central counterargument against agent-causation. Namely, the argument that we cannot regard reasons as the causes for an agent-caused action, and that defenders of non-event-casual theories are not able to account for reasons the same way event-causal theories can (Clarke, Capes, Swenson, 2021, 2.1). This argument is based on the idea that agent-caused actions cannot be caused by the agent recognizing the reasons for the action. This is because the recognition of reasons would be an event, which would mean that the action was event-caused rather than agent-caused. (ibid, 3.0).[15]

O'Connor believes that while reasons can explain actions, this does not mean they directly produce the actions they explain (O'Connor, 2000, p95). This idea can be best understood by contrasting event-causal and agent-causal libertarianism. From an event-causal view, the reasons behind performing an action need to, to some extent, be part of bringing that action about if we are to consider it as an explanation for why the action happened. From an agent-causal view, the reasons behind the action need instead only explain why the agent acted as they did, but without the reason directly causing the action. Both viewpoints uncontroversially agree that the reasons behind an action play a part in its performance. However, event-causal libertarians can regard reasons as causes, while agent-causal libertarians cannot. With this in mind, let us look at how

---

[15] O'Connor adds to this point by noting that due to the prominence of compatibilism in contemporary western philosophy, many philosophers believe that reasons can explain an action, but only in so far as they help in bringing the action about (O'Connor, 2000, p85).

O'Connor attempts to defend agent-causal libertarianism by pointing out how desires that lead to intentions are sufficient to explain an action without appealing to the recognition of reasons as the cause of an action.

> 1. prior to the action, the agent had a desire that O and believed that by so acting he would satisfy (or contribute to satisfying) that desire;

> 2. the agent's action was initiated (in part) by his own self-determining causal activity, the event component of which is the-coming-to-be-of-an-action-triggering-intention-to-so-act-here-and-now-to-satisfy-O;

> 3. concurrent with this action, he continued to desire that O and intended of this action that it satisfy (or contribute to satisfying) that desire; and

> 4. the concurrent intention was a direct causal consequence (intuitively a continuation) of the action-triggering intention brought about by the agent, and it causally sustained the completion of the action (O'Connor, 2000, p86)

In other words, a person has a desire and believes some action will help them realize that desire (1). Due to having this desire (1), the person starts to perform an action that, at least in some respects, is caused by themselves (2). The person continues to hold on to the original desire (1) and performs some action that they believe will satisfy this desire (3). During the action, the person continues to have an intention which leads to the performance of the action (4), and this intention was caused by the original intention to perform this action to satisfy their desire. Notably, on O'Connor's account, agents do not cause free actions. Instead, a free decision is a causally complex event where the agent causes themselves to have a certain intention, and any free overt action is partly constituted by a complex event of this kind (Clarke, Capes, Swenson, 3.1, 2021).

This way, O'Connor builds upon Reid's ideas, as they share the view that we can exert a form of mental effort to cause actions without this causation reducing to event-causation. O'Connor also explains the role of reasons, arguing that reasons merely have an explanatory role rather than a causal role in actions of this kind. With that, O'Connor, Chisholm and Reid have given us a feel

for the plausibility of agent-causality in so far as it can be explained in non-event-causal terms and be reconciled with a realist account of causation.

Lastly, a key argument against both agent-causal and event-causal libertarianism is the luck objection. This objection concerns the problem of how indeterminism seems to reduce control by introducing luck (Clarke, Capes, Swenson, 2021, 2.0). In other words, if it is indeterministic what action one performs, how can it not also be said to be a matter of luck what action one performs? Some philosophers have argued that agent-caused actions are a matter of luck by pointing out that opting to exercise agent-causal powers at a specific time might have resulted in a different outcome in a different possible world (Mele, 2006, p70). Other philosophers have contended that agent-caused actions are not lucky. Clarke pointed out in 1993 that previous contemporary accounts of agent-causation, though they had been rejected at the time of writing, held that actions that are agent-caused are not matters of chance or accidents (Clarke, 1993, p193). Griffith also shows support for this idea by questioning how a person who does something freely for her own reasons can be said to be lucky (Griffith, 2010, p52). Pereboom has also stated that he does not believe that agent-causal libertarianism falls to the luck objection, as opposed to event-causal libertarianism, the agent still plays a part in choosing what action they bring about after a set of antecedent events if agent-causal libertarianism holds true (Pereboom, 2007, p110-111). In other words, Pereboom agrees with Mele that the agent might have acted differently in a different world but contends that whichever action they perform is not a matter of luck, as it is still the agent as a substance that causes the action.

In light of this, I find agent-causal libertarianism to be more robust against the luck objection than event-causal libertarianism. I say this because, on an agent-causal view, the agent, in one way or another, served as the cause of an event. While not fully uncontroversial, I therefore agree with Pereboom's sentiment that being the originator of an action also makes an agent responsible for the action. Any agent-caused action must then be indeterminate, but if the action was caused by the agent intentionally, calling it a mere matter of luck seems a mischaracterization of why the action was caused. Put another way, calling an action a matter of luck implies that the action was random. However, if the action had an intention behind it and was self-caused due to an intention and a desire, it seems misinformed to call some such action random. Therefore, I find agent-causal libertarianism more robust against the luck objection than

event-causal libertarianism, as the event-causal libertarianism cannot appeal to a first cause in the same manner. However, despite these abovementioned strengths, I find that agent-causation leaves much to be desired in terms of providing a sufficiently comprehensive account of free will that is reconcilable with modern science.

2.4 The issues of Agent-Causal Libertarianism

In this section, I will lay out counterarguments against agent-causal libertarianism, arguing that despite the abovementioned strengths, I ultimately do not find it worthwhile to endorse this view. To me, the most glaring issue with agent-causation is how the position's mystic nature makes it implausible and difficult to grasp. A similar sentiment is expressed by Peter van Inwagen, who states he does not believe he understands agent-causation, and that it is a mystery that a thing, as opposed to a change in a thing, could cause a change (van Inwagen, 2015, p280). This mysteriousness is a crucial point in the discussion on agent-causation, with Kane pointing out that even defenders of the position have pointed out the position's unusual and mysterious nature (Kane, 2005, p45, p47).[16]

This issue can be expressed in a multitude of ways. For example, in the previous section, I looked at how O'Connor, building on the ideas of Reid and Chisholm, provided an account of agent-causation in which agents do not directly cause free actions. Rather, by causing themselves to have an intention and thereby bringing actions about, agents are free in causing a part of a complex event that is a component of a free action. However, despite accounting for some counterarguments against agent-causal libertarianism, O'Connor also becomes prone to new ones when providing this account. Notably, Clarke and colleagues point out how O'Connor's account commits him to an implausible view concerning the explanation of causally complex events (Clarke, Capes, Swenson, 2021, 3.1). To illustrate this point, they use an analogy of a different type of causally complex event and how a state seemingly cannot be cited as the explanation for the causation of the event.

> Suppose that a flash of lightning has caused a brush fire. A drought, let us say, had left
> the brush dry, and had this not been so, the lightning flash would not have caused the fire,

---

[16] O'Connor has also said that he does not throw accusations of incoherence around lightly, as he is a believer in agent-causation himself (O'Connor, 2000, p97).

or at least would have been less likely to do so. Now suppose that the dryness of the brush is cited as an explanation of the flash's causing the fire. Does the truth of the proffered explanation hinge on whether the fire has a certain feature, and can the explanation be true even if the brush's being dry did not cause either the flash, the fire, or the flash's causing the fire? Apparently not. (Clarke, Capes, Swenson, 2021, 3.1)

The point made here is that using the dry state of the bush as an explanation for the flash causing the brush fire is vastly different from how we ordinarily think about causation. This is because we cannot cite the dry state of the bush as the explanation for the flash, fire, or the fire causing the flash, given that the dry state did not cause any one of these. Whether the dryness of the brush explains the flash causing the fire also does not seem to depend on whether the fire has a certain feature.

I will now explain how this analogue is relevant to O'Connor's view. To reiterate, his reason-explanation of free actions is as follows:

1. prior to the action, the agent had a desire that O and believed that by so acting he would satisfy (or contribute to satisfying) that desire;

2. the agent's action was initiated (in part) by his own self-determining causal activity, the event component of which is the-coming-to-be-of-an-action-triggering-intention-to-so-act-here-and-now-to-satisfy-O;

3. concurrent with this action, he continued to desire that O and intended of this action that it satisfy (or contribute to satisfying) that desire; and

4. the concurrent intention was a direct causal consequence (intuitively a continuation) of the action-triggering intention brought about by the agent, and it causally sustained the completion of the action (O'Connor, 2000, p86)

(1) lays out the state of an agent prior to an action. (2) details how the action is initiated, namely, desiring O and believing that some act would aid the agent in satisfying O. This is the first step in how we can agent-cause parts of a causally complex event. However, just like the dryness of the brush, (1) is a state the agent is in before the causation in question happens. Just like the flash might not have caused the fire if the bush was not dry, a causally complex event might not have

occurred without (1). However, this does not mean that (1) can explain how, on O'Connor's view, an action-triggering-intention casually sustains the completion of an action in step (4). The claim that the initial state of an agent prior to the causally complex event explains how an intention caused an action is, therefore, akin to citing the dryness of the brush as the explanation for the flash causing the fire.

Meanwhile, O'Connor's account hinges on there being an agent-causal aspect to a causally complex event. This agent-causal part, the causing of an intention, can explain how the intention to raise a hand, the raising of the hand itself, and how this intention causes the raising of the hand. However, when we apply the same logic to the case of the brush fire, we can see how O'Connor's account vastly differs from how we more commonly speak and think about causation. Put another way, O'Connor merely claims that rather than causing a free action, the agent is causing themselves to have a certain intention, which is part of a complex event which, in turn, is part of any free action. However, if the state prior to this intention is cited as the explanation for why an agent performs some action such as raising their hand, this explanation seems vastly different from how we ordinarily think about causation, given that this state did not cause the raising of the hand. Neither the raising of the hand or the brush fire would be possible, or at the very least as likely, without the state of dryness or the state of desire. But, if we cannot reasonably cite the dryness as an explanation for what caused the brush fire, given that it did not cause the fire, why should we think differently about the explanatory capabilities of a state in a causally complex event involving an agent? Despite being able to bypass the issue of citing the recognition of reasons as the cause for an action, the difficulty of answering this question shows that when more thoroughly examined O'Connor provides a problematic account of causation.

Another issue is that O'Connor's view also needs to account for whether an agent could avoid performing some action when they have a desire and form an intention to perform some action to fulfil that desire. This stems from an issue pointed out by David Widerker; the claim that A could agent-cause event B, also means that A could have refrained from agent-causing event B (Widerker, 2005, p93).[17] In other words, even if we grant O'Connor that his account of agent-

---

[17] Specifically, Widerker is arguing against a stance he calls strong agent-causal libertarianism. O'Connor's account of agent-causation is such an account, and Widerker lays out the defining trait of this type of view as follows: if there is such a thing as agent-causation, then one can show that, as a matter of conceptual necessity, agent-causal acts are within the agent's control (Widerker, 2005, p87).

causation in complex events is correct, the issue remains of whether these complex events result in free actions. Because of this, Widerker also argues that defenders of agent-causation cannot prove the reality of free actions, even if we assume agent-causation to hold true (ibid, p95). Therefore, it is difficult to see how O'Connor's account of agent-caused actions truly explains how an agent can be said to cause an event. Even if we concede that agent-causation of this kind is possible for the sake of argument, it is difficult to see how agent-caused actions are ultimately in the agent's control. On O'Connor's view, there is no agent-causation that has yet occurred when we form a desire and believe some action will satisfy the desire. However, this desire will unquestionably shape any intention and action that follow it, but given that we did not cause the desire or belief that some action would satisfy it, we also do not seem to be in control of what intention we form, regardless of whether this intention was agent-caused.

A similar point can be made by considering Chisholm's view on how we can understand our own casual tendencies. As previously mentioned, Chisholm, in echoing Reid, has pointed out that it can never be predicted with absolute certainty what an agent-caused action will be and that it is through understanding our own causal tendencies that we can understand causation at all (Chisholm, 1964, p11-12). However, if understanding "our own casual tendencies" is what it would take to understand causation, I would be even less inclined to believe that we can choose not to agent-cause certain events. If we can understand what events we might cause better through experience, this view seems no different from scientific observations of how causation occurs among events. It would take experience for me to know that strong storms have a casual tendency to blow trees over. Similarly, it would take experience for me to know that I have a causal tendency to drink coffee in the morning. In turn, it can be questioned whether our actions can be called agent-caused. Especially so because an agent-caused action would require the ability to refrain from performing the action, and if we cannot predict but merely understand our own casual tendencies through observation, this view seems more in line with determinism than with libertarianism.

Another major concern with agent-causal libertarianism is the question of whether the view can be reconciled with contemporary science. Potter and Mitchell have pointed out how this view is often seen as mysterious and does not fit with the widespread physicalist ideas in modern science and philosophy (Potter, Mitchell, 2022). Pereboom, despite conceding that agent-causation might

be conceptually possible, does not find the likelihood of us having agent-causal powers impressive (Pereboom, 2007, p197). He also writes that us having agent-causal powers is not credible given our best physical theories (ibid, p85). Potter, Mitchell, and Pereboom's critiques are also closely tied to what Robert Kane calls the intelligibility problem (Kane, 2007) or the intelligibility question (Kane, 2002, 2014). This problem concerns how we can have free will if indeterminism holds true, rather than indeterminism being a matter of chance we have no control over. In discussing what modern libertarians must do in the face of contemporary science, Kane has the following to say about this problem:

> But they must also show (ii) that a libertarian free will requiring indeterminism can be made intelligible and how, if at all, such a free will can be reconciled with modern scientific views of the cosmos and of human beings (call this "The Intelligibility Problem") (Kane, 2007, p9).

In other words, the intelligibility problem is a a question that needs to be answered by libertarian accounts if they are to be made intelligible, and it has become increasingly difficult to do so due to modern advancements in science.[18]

One reason the intelligibility question is especially relevant today is due to the physicalist ideas that the mechanical parts of humans are what cause us to hold causal powers.

> ..causation is attributed to a specific subset of components *within* the system—it is a particular 'part' (or set of parts) that causally determines the action of the 'whole'. Thus, the whole does not have any causal power in virtue of being a whole because causation is entirely localised to (a set of) components with smaller spatial and temporal dimensions than the whole. (Porter, Mitchell, 2022, 1).

In other words, causality happens at the level of the smallest components rather than at the highest level, and therefore, we cannot be said to cause our own actions directly. This is arguably a problem for any type of libertarianism, but Kane finds this problem particularly detrimental for

---

[18] It is worth noting that any libertarian theory, be it agent or event-causal, still relies on there being genuine indeterminism in nature (Kane, 2005, p133). Despite Kane having to answer the intelligibility problem himself as a libertarian, he argues that libertarianism is difficult to defend on an agent-causal account. This is because agent-causation is an extra-factor strategy that might not be reconcilable with how indeterminism occurs in nature, whereas his event-causal libertarianism does not face this problem.

agent-causal accounts. This is because the abovementioned physicalist idea does not seem reconcilable with what Kane calls extra-factor strategies, which attempt to explain human actions through a unique form of causation that differs from and cannot be explained in terms of our ordinary idea of causation (Kane, 2007, p25). Historically, some of the most noteworthy explanations for human causation have been extra-factor strategies, such as Kant's noumenal selves and Descartes' Immaterial egos & Dualism (Kane, 2002, p415, 2005, p40-44). As for the modern free will debate, Kane points out that agent-causation is the most common extra-factor strategy and that he wishes to find other explanations for how to make sense of libertarian free will (Kane, 2002, p415-p416).

I find that the intelligibility problem is especially detrimental for older works on agent-causation, particularly Reid's and Chisholm's. The claim that agent-causation occurs in our brains in one way or another has been difficult to prove, both due to the elusive nature of agent-causation and due to our lack of understanding of our brains. However, as scientific discoveries make it harder to see how Reid's and Chisholm's claims about brain states could hold true, their accounts have become less elusive but all the more implausible. This is, of course, an issue that any agent-causal view must account for, but one that is especially glaring for works that predate modern science and cite brain activities in explaining agent-causation.

Timothy O'Connor has responded to Kane regarding the concept of intelligibility. He argues that the advances in modern science do not justify disregarding agent-causation. Instead, O'Connor claims we should try to reexamine our understanding of agent-causation, given modern scientific discoveries. (O'Connor, 1996, p156-157).[19] In my view, this puts agent-causation in a weak light. If the project of agent-causal libertarians is to reexamine the view to make it compatible with modern science when the view already faces a multitude of strong counterarguments, I see no reason at the current moment to believe that we have agent-causal powers. For all we know, we might have agent-causal powers, but I argue that defenders of the view have yet to provide a plausible account of why this would be the case, and in light of modern science, the view stands as weak as ever.

---

[19] O'Connor is not trying to reconcile how agent-causation can fit with modern scientific discoveries (O'Connor, 1996, p143). Instead, he is pointing out that we should not be so quick to disregard agent-causal libertarianism as a position because we are unable to explain it in light of new findings and that incompatibilists still ought to believe we have agent-causal powers (ibid, p144).

I also find that the abovementioned issues make the robustness of agent-causation against the luck objection insignificant. My argument is similar to how Pereboom argues that we most likely do not have moral responsibility due to the lack of evidence that we have agent-causal powers.

1. Event-causal libertarianism could not conceptually hold true, but agent-causal libertarianism could conceptually hold true (Pereboom, 2001, p40-59)
2. Agent causation is not credible given our best physical theories (Pereboom, 2007, p85, p112)
3. Therefore, given that we have no evidence of the indeterminacy required for agent-causal libertarianism to hold true, we also have no evidence that we are morally responsible (Pereboom, 2001, xviii)

Similarly, I find that agent-causation would escape the luck objection if it were to hold true, but given that we most likely lack agent-causal powers, we also lack a robust defense against the luck objection. If we could agent-cause our actions, I have argued that it would be misinformed to call our actions lucky. However, this is nevertheless insignificant if we do not have agent-causal powers. Therefore, while showing how indeterminate actions can be morally responsible is a potential strength of agent-causation, it is a strength that requires the view to be sufficiently plausible to begin with. For the reasons outlined above, I argue that it is not.

3. Event-Causal Libertarianism

3.1 Self-Forming Actions & Parallel Processing

In the previous chapter, I discussed agent-causal libertarianism and why I do not find this position worth endorsing. In this chapter I will examine a different form of libertarianism in Robert Kane's event-causal libertarianism. I will begin by outlining his account. Then, I will show how his account circumvents many of the problems agent-causal libertarianism runs into. Lastly, I will argue for why I ultimately do not endorse this position either.[20]

As mentioned in Chapter 2, Kane is skeptical of agent-causal libertarianism and claims that agent-causation creates further mysteries (Kane, 2007, p25). He instead endorses a form of event-causal libertarianism.[21] This means that he believes free will is possible, incompatibilism holds true, and that free actions are caused by states or events. Therefore, like agent-causal libertarians, event-causal libertarians view indeterminacy as a prerequisite for moral responsibility (Pereboom, 2007, p101). This is because in endorsing incompatibilism, defenders of the view claim that determinism would preclude moral responsibility. A key strength of event-causal libertarianism is that the view avoids the challenge of accounting for the mysticism that agent-causation can bring with it of explaining how something other than an event can cause something to happen. On the other hand, defenders of the position are faced with the difficulty of accounting for how we can be held morally responsible for actions that are indetermined and how we can be said to have free will if our actions are the result of events.

---

[20] There are other forms of event-causal libertarianism than Kane's that can be worth considering. Kane's view is a centered account. Centered accounts claim that some free actions must be nondeterministically caused by their immediate causal antecedents and thus require indeterminism "centered" on the production of free actions (Clarke, Capes, Swenson, 2021, 2.2). Deliberative indeterminism, on the other hand, is a view that claims that indeterminism can instead occur in the earlier stages that lead up to an action without diminishing the agent's control (ibid, 2.1). Problematically, this view has to claim that agents can sometimes have free will in decisions that are not up to the agent (ibid, 2.1). Not only do I find this implausible, but this idea differs significantly from how free will is otherwise discussed in this thesis. I have instead chosen to focus on Kane's centered account as it provides an in-depth explanation that attempts to account for many common arguments against libertarianism without running into this problem. Therefore, I find it sufficient to address Kane's view to highlight my main issues with event-causal libertarianism, especially so because several of my counterarguments against Kane could also be raised against other forms of event-casual libertarianism.

[21] Note that Kane has gone on to reject the "event-causal" label (Clarke, Capes, Swenson, 2021, 2.2). His stance is, however, commonly referred to as event-causal libertarianism by others (Pereboom, 2007, p103-p106, Franklin, 2014, p414, 417, Clarke, Capes, Swenson, 2021, 2.2).

A cornerstone of Kane's event-causal libertarianism is a concept he calls self-forming actions, henceforth SFAs. As SFAs are a nuanced concept I will proceed by outlining this concept in detail. According to Kane, SFAs occur in situations where we are faced with a decision on what to do and our options have competing reasons or motives. (Kane, 2007, p26-29). When faced with decisions of this kind, we have a desire to individually choose each of the options, but we can only ever choose one of them. Kane holds that SFAs happen at difficult times when we have competing visions of who we should be or become (ibid, p26). More specifically, Kane believes there are at least 6 types of willing's that can result in SFAs:

1. Moral choices and decisions.

2. Prudential choices and decisions.

3. Practical choices and judgments.

4. Efforts of will sustaining purpose.

5. Attentional efforts which are directed at self-modification.

6. Changes of intention in action. (Kane, 1996, p125, my formatting and wording)

For example, in the case of a moral choice where an agent makes a SFA, an agent could be torn between a solidary choice and a self-serving choice. Then, the agent who must decide between these two options cannot choose both, and in committing to one of the options, they form their character. In this example, the agent would be forming their character to become more selfish or solidary. Kane also regards this inner conflict and the necessity of landing on a choice as indetermined. Therefore, the agent could choose either of the two options. Furthermore, Kane regards whichever choice we make as a meaningful path that is consistent with our pasts, and despite us not conclusively knowing which option to choose over another, we have our reasons for choosing that option in particular (Kane, 2007, p41-42).

It is also important to note that Kane does not hold that in order for an action of this kind to be self-forming, we would need to choose an option we deem to be more virtuous or that we need to have a triumph of the will where we overcome our temptations. Rather, it is the indeterminacy in our efforts that result in determinate choices during a self-forming action that makes them self-forming. Kane exemplifies this idea by encouraging us to imagine a woman torn between

making an egotistical, career-furthering choice to go on to a meeting and an altruistic choice of staying to save another person and thus not making the meeting in time:

> She has to make an effort of the will to overcome the temptation to go on. If she overcomes this temptation, it will be the result of her effort, but if she fails, it will be because she did not allow her effort to succeed. And this is due to the fact that, while she willed to overcome temptation, she also willed to fail, for quite different and incommensurable reasons. When we, like the woman, decide in such circumstances, and the indeterminate efforts we are making become determinate choices, we make one set of competing reasons or motives prevail over the others then and there by deciding (Kane, 2007, p26)

In other words, it is not in virtue of the outcome decided upon but rather due to how that outcome was eventually decided upon through an indeterministic process where one set of values triumphs over another that some of our actions can be self-forming.

The idea of wanting to do each of two competing things at once seems uncontroversial and no doubt something all of us experience regularly. However, Kane makes the somewhat more controversial claim that, in these cases, the agent is *attempting* to make both choices (Kane, 2007, p30). In other words, there is a competing effort to do each of two things at once despite only ever being able to do one. Kane's explanation for how this dual effort of trying to make both choices is possible is due to parallel processing (Kane, 2002, p419, 2007, p28). To understand the concept of parallel processing and how it relates to SFAs, we must first consider how Kane believes indeterminacy comes about at the neuronal level.

> There is tension and uncertainty in our minds about what to do at such times, I suggest, that is reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium - in short, a kind of "stirring up of chaos" in the brain that makes it sensitive to micro-indeterminacies at the neuronal level. The uncertainty and inner tension we feel at such soul-searching moments of self-formation is thus reflected in the indeterminacy of our neural processes themselves. (Kane, 2007, p26)

In situations that result in SFAs, there is a physical reaction in our brains that results in instability. This instability serves as an explanation for why we feel torn between two options in

situations that result in SFAs. The choice feels indeterminate to us because at the neural level, where this uncertainty stems from, the choice is undetermined. In other words, when brain instability of the abovementioned kind occurs, two conflicting options would physically influence and compete with one another at the neuronal level.

To understand parallel processing, Kane urges us to imagine two crossing recurrent neural networks that can influence the other, which in turn results in chaos (Kane, 2002, p418, 2007, p28). Chaos is a much debated concept in scientific literature (Basar, 1990, Diaz & Colleagues, 2015). Christine Skarda & Walter Freeman write that some views on chaos have shifted from chaos being regarded as undesirable noise to being seen as beneficial to us and serving as a way for us to make sense of the world around us (Skarda & Freeman, 1987, 1990). Chaos is foundational for Kane's event-causal libertarianism, and part of why he believes that trying to do two competing things can, in some cases, be rational (Kane, 2007, p34-35). In other words, chaos is not to be mistaken for irrationality, and he holds a view of the kind outlined by Skarda and Freeman, where chaos is instead something we use to make sense of the world. Kane illustrates this point by comparing the solving of this chaos to that of solving a mathematical puzzle, as it is indeterministic whether you will succeed, and up until the moment the conflict gets resolved, there is a neural noise that is distracting you from completing it (Kane, 2002, p418).[22] I interpret this as an analogue to trying to make a choice where up until the resolution (committing to attempting an action or, analogously, solving the puzzle), the chaos and parallel processing at the neural level are indeterministic concerning whether you will succeed, or what the outcome of your deliberation will be.

Kane bases his view on how indeterminacy happens at the neural level in this manner and relates this idea to how agents can be held morally responsible. If an agent went through the deliberation and effort of trying to make both choices or made an attempt to solve the mathematical puzzle, the agent is responsible regardless of the outcome, as despite the outcome being indeterministic, it was nevertheless done intentionally and on purpose. (Kane, 2007, p29). As the action was done with intention and purpose, it also follows for Kane that the agent can be held responsible for the

---

[22] While Kane regards chaos of this kind as part of an indeterministic process, note that whether this neural "stirring up of chaos" truly is indeterministic or deterministic has been disputed (Vargas, 2007, p135, p143, Kane 2007, p182).

action. This is because regardless of what the agent ends up doing, the agent was attempting to do just that.

This way, Kane provides an answer to the intelligibility question and the luck objection. To reiterate, the intelligibility question concerns whether we can be held morally responsible if either indeterminism or determinism holds true, and the luck objection points out how indeterministic events seem to be a matter of chance. Kane answers that we can be held morally responsible for indeterministic actions as actions are done with purpose and intentionality and, therefore, not mere matters of chance despite being indeterministic. Furthermore, this indeterminism is also why, for Kane, some actions are more than merely matters of luck, as the neural-level indeterminism is not to be regarded as random but rather as the reason actions can be said to be free to begin with.

However, there is still more to be said about Kane's idea of SFAs as parallel processing alone cannot solely answer many of the objections that event-causal libertarianism commonly faces. In order to account for common counterarguments, Kane has outlined a set of conditions that need to be met for an SFA to occur. These conditions lay the groundwork for how people can be said to be free by shaping themselves and their future by deciding in the moment, and when all these conditions are met, we have what Kane calls plural voluntary control (Kane, 2007, p30). The conditions for plural voluntary control all concern what type of ability we need to have when performing SFAs, and they are as follows:

> (1) An ability to
>
> (2) bring about
>
> (3) at a specific time
>
> (4) whichever of the options they will or want,
>
> (5) for the reasons they will to do so,
>
> (6) on purpose or intentionally rather than accidentally, by mistake or merely by chance (as when by mistake I press the wrong button on the coffee machine), hence
>
> (7) voluntarily (in accordance with their wills rather than against them),

(8) as a result of their efforts, if effort should be required,

(9) without being coerced or compelled, or

(10) otherwise controlled or forced to choose one way or the other by some other agent or mechanism. (Kane 2002, p430)

Kane answers defender of agent-causation when he argues that the conditions for plural voluntary control can be met without appealing to non-event causation even if the choice options in an SFA are undetermined (Kane 2002, p430).

In part, this is expressed in condition (5) for plural voluntary control: The ability to bring about an SFA "for the reasons they will to do so". In chapter 2 I discussed how O'Connor claimed that the recognition of reasons cannot cause us to act on an agent-causal view. This was because if the recognition of a reason leads to an action this recognition would be an event and thus make the action event caused. Instead, reasons had to be accounted for as they clearly play a role in bringing actions about, but they could not be given full casual power on O'Connor's view. This, in my view, made O'Connor's account less robust. With condition (5), Kane can account for what role reasons play in free actions while not being prone to the same counterarguments as O'Connor.

As previously mentioned, Kane also responds to the luck objection, which is further expressed in (6).[23] Furthermore, Kane responds to how some compatibilists could opt to argue against his view with conditions (4), (9) and (10).[24] Most importantly, Kane believes that performing SFAs in certain situations where we fulfil all the conditions required for plural voluntary control results

---

[23] The luck objection, simply put, objects that we can have moral responsibility for undetermined actions (Pereboom, 2007, p102). This is because if an action is undetermined, how can it not be said to merely be a result of chance and therefore also a matter of luck. With condition (6), the ability to do an SFA "on purpose or intentionally rather than accidentally, by mistake or merely by chance (as when by mistake I press the wrong button on the coffee machine)" Kane acknowledges the threat of the luck objection. However, with (6) he makes the claim the objection can be circumvented due to us performing actions intentionally and with purpose when an SFA happens.

[24] Some compatibilists hold that we can be morally responsible without alternate possibilities and that moral responsibility is compatible with determinism (Mandik, 2013, p178). Kane denies this in condition (4), the ability to do "whichever of the options they will or want". Some compatibilists have argued this point by showing that a person who is coerced or forced to do something and thus does not have alternate possibilities can nevertheless be held morally responsible (McKenna, Coates, 2024, 3.2). Kane denies this with conditions (9) and (10), the ability to act "without being coerced or compelled, or otherwise controlled or forced to choose one way or the other by some other agent or mechanism".

in us having ultimate responsibility. Ultimate responsibility is a core concept for Kane's view, and I will, therefore, dedicate the next section to thoroughly introduce it.

3.2 Ultimate Responsibility & Austin-Style Examples

In this section, I will explain Kane's view on ultimate responsibility and illuminate why he finds it a crucial concept for his account of free will. For us to have ultimate responsibility, henceforth UR, Kane writes that the following condition must be met:

> Free will also seems to require that the sources or origins of our actions lie "in us" rather than in something else (such as the decrees of fate, the foreordaining acts of God, or antecedent causes and laws of nature) outside us and beyond our control. (Kane, 2007, p14)

For Kane, having UR is crucially important for being free and morally responsible (Kane, 2007, p14). To illustrate why, I will outline how we can be said to have UR on Kane's view. I will also show how having UR provides an answer to how we can have free instead of our actions merely being a result of chance if indeterminism holds true. According to Kane, people are responsible for their SFAs when they succeed in performing the action they intended despite not knowing whether they have made the right choice (Kane, 2007, p37-42). Thus, we have moral responsibility when successfully performing SFAs as we endorse certain values and shape our dispositions when we perform them. In turn, SFAs makes us responsible for subsequent actions, even if the following actions are not self-forming themselves. This is because we are now responsible for what type of people we have shaped ourselves into. Our disposition makes us choose and endorse certain actions, but we are ultimately responsible for our disposition as we shaped it through performing SFAs.[25]

In Chapter 1, I discussed the principle of alternate possibilities (PAP). To reiterate, this principle states that we need to be able to choose otherwise to be held morally responsible, and due to how

---

[25] In chapter 2, I discussed how Aristotle made the claim that a person must be responsible for forming their character if they are to be held responsible for the acts that flow out of their character. Kane argues that this claim is compatible with UR and that we would only need alternate possibilities sometimes in the form of SFAs (Kane, 2007, p14). There is also a further parallel between Kane's and Aristotle's views. According to Aristotle, one needs to perform virtuous acts to become virtuous (Kraut, 2022, 4), which resembles Kane's view of how we become responsible for our dispositions through performing SFAs.

intuitive the principle seems, it has been a major point of discussion in the free will discourse. However, Kane claims that UR is even more important to the free will debate than PAP (Kane, 2007, p14). He also writes that to understand the connection between UR and PAP, we must note that PAP might be necessary for free will, but it is not sufficient for free will even if the alternate possibilities are undetermined (ibid, p17). To illustrate this point, he uses thought experiments known as Austin-style examples. Austin's original example concerns an attempted golf putt that he misses due to a nervous twitch in his arm:

> Consider the case where I miss a very short putt and kick myself because I could have holed it. It is not that I should have holed it if I had tried: I did try, and missed. It is not that I should have holed it if conditions had been different: that might of course be so, but I am talking about conditions as they precisely were, and asserting that I could have holed it (Austin, 1956, p218)

Examples of this kind have been used to showcase how agents can be said to be able to do otherwise, even if an event is undetermined (Kane, 2002, p409). Unlike Frankfurt cases, which attempt to show that we can be held morally responsible if determinism holds true, Kane uses Austin-Style examples to show that we can be held morally responsible, as opposed to being victims of chance, if indeterminism holds true.

Austin may or may not have holed the put, making the action indetermined. However, the circumstances are also so that it would be strange to say that it was solely a question of luck whether he holed the put. After all, his capacity and abilities still played a major role in whether he succeeded. Kane points out how Austin believes that this is a scenario in which he "could have done otherwise", in virtue of the event being undetermined, and as Austin could have been able to make the putt as he was a good putter and had succeeded in making similar putts in the past (Kane, 2007, p17). This way, the examples provide one type of answer to how we can be held morally responsible for an undetermined action. If we agree with Kane that alternate possibilities is a necessary requirement for moral responsibility, and Austin's examples show how we can have alternate possibilities when performing undetermined actions, we now have an account of how we can be morally responsible even if indeterminism holds true. If this is the case, this view also circumvents the luck objection, as the outcome of undetermined actions of this kind depends on ability instead of mere chance.

However, this also raises a further question. Even if the actions in Austin-style examples are undetermined while keeping the option of being able to do otherwise open, the question arises of whether these cases concern truly free human actions. I say this because even if missing the putt and holing the putt were both possible outcomes in Austin's example, missing the putt due to a nervous twitch can hardly be said to be a purposeful action. Austin could have holed the putt, but he did not intend to miss the putt and, therefore, also seemingly did not freely miss the putt. From this, the question arises of whether the outcome of missing the put is a result of a free action and, for our present purposes, whether the outcome cannot still be called a matter of chance. This is why Kane's point (6) for plural voluntary control, an ability to do something "on purpose or intentionally rather than accidentally, by mistake or merely by chance" is so crucial. (6) denies that Austin-style examples are scenarios where we have plural voluntary control, and (6) showcases how actions of this kind are not self-forming despite being undetermined and accounting for PAP. Instead, Kane believes that for us to have free will as a result of performing SFAs, we overcome obstacles and hinderances in our minds (Kane, 2002, p425). In doing so, we endorse one option over another, and despite the action being undetermined until the moment we choose, the action nevertheless differs from Austin-style examples in that it is voluntary.

This point can also be understood from a different angle. Kane states that SFAs of this kind are not required by an agent's past but are consistent with the agent's past (Kane, 2007, p41). In other words, the action is indetermined as it is not determined by the past, and the SFA also aligns with the agent's value system, which has been shaped by previous SFAs. Therefore, SFAs are not merely a result of chance or luck, though they would have been if they were the result of a nervous twitch rather than being caused by a voluntary effort. We can see this contrast between Austin-style examples and SFAs more clearly when we look at how Kane describes the way a person who commits to a choice that leads to an SFA could explain their rationale.

> I did have good reasons for choosing as I did, which I'm willing to endorse and take responsibility for. If they were not sufficient or conclusive reasons, that's because, like the heroine of the novel, I was not a fully formed person before I chose (Kane, 2002, p425, 2005, p145)

I find that this point made by Kane' can be illustrated by applying his account of free will to Austin-style example. It makes no sense to speak of Austin having conflicting wants of missing

the putt and hitting the putt, and in deciding on either option, he attempts to hit or miss it. Instead, the outcome of the put comes about unintentionality, which produces a result that, as Kane would likely put it, Austin does not endorse. This way, we can then see how little sense it makes to speak of having free will purely in virtue of actions being indetermined.

As Kane's above-quoted statement illuminates, an SFA cannot produce anything but an outcome that the agent endorses. This holds true regardless of the motive for the action. Altruistic and self-serving choices are endorsed in the same way when an SFA occurs. As for the issue posed by the luck objection, missing the putt is unlucky in the sense that the twitch that caused the miss was unintended and undetermined. On the other hand, choosing an option during an SFA is not lucky nor unlucky, as despite being indetermined, the resulting action was done intentionally, purposefully and voluntarily. Suppose we accept Kane's view that we can have plural voluntary control and perform SFAs, and that this, in turn, gives us ultimate responsibility. In that case, we seem to have circumvented the luck objection, as Kane has shown why SFAs are not just a matter of chance. We can also see why he claims UR is more important than PAP, as PAP is not sufficient for free will, while it is in virtue of having UR that we can be said to have free will at all.

By claiming that we have UR due to performing SFAs, Kane also answers the argument that we need to be responsible for our disposition, which makes us act the way we do if we are to be held responsible at all. As mentioned in chapter 2, this argument was raised by Aristotle, and it continues to be raised to this day. Notably, Chisholm has expressed this problem by saying that if we are to be held responsible for the beliefs and desires we have, it must at some point in time have been in our power to acquire or not to acquire these beliefs or desires (Chisholm, 1964, p6). Kane also points out how Aristotle expressed a similar sentiment in saying that a man must have at some point been responsible for forming a wicked character if he is to be held responsible for the wicked acts that flow from his character (Kane, 2007, p14).

If we have UR, this would seem to account for the issue of responsibility and character formation raised by Aristotle and Chisholm. This is because SFA and ultimate responsibility account for how we can be held morally responsible not just in virtue of who we are (disposition) but also in virtue of how we form this disposition through what we do (endorsement). In other words, by forming our character through SFAs, we can, according to Kane, be ultimately and morally

responsible for who we are. In turn, this also makes us morally responsible for subsequent actions we perform due to who we are, whether these subsequent actions are themselves self-forming or not. Therefore, if Kane is correct, this would mean that we can be responsible for forming a wicked character without us needing to be uncaused causes of ourselves. Kane's event-causal libertarianism can, therefore, serve as an attractive alternative to incompatibilists who hold that we have free will but reject that agent-causation holds true.

3.3 The Issues of Chaos, Effort & Intention

In the last two sections, I outlined Kane's event-causal libertarian account and showed some of the strengths of his position. I will now proceed by critiquing his account. In this section, I will specifically critique how Kane's view of chaos, effort, and intention is lacking in various aspects. In order to make this claim, I will first need to lay out what I find to be the most central issue with Kane's view. Namely, Kane's event-causal libertarianism requires indeterminism to work in a specific way. Even if this indeterminism is present in this specific way, such as when all the conditions for having plural voluntary control are met, I argue that the luck objection has not been circumvented. I say this because regardless of whether SFAs can be understood as intentional and voluntary as opposed to Austin-Style examples, SFAs nevertheless cannot account for how it is not a matter of luck which of the conflicting options the agent is torn between they ultimately end up endorsing. I concede that SFAs are superior to Austin-style examples in accounting for how we could be held morally responsible for indeterministic actions. However, I do not believe that it follows from this that we have circumvented the issue of luck in the manner that Kane claims.

At first glance, choosing an altruistic option over a self-serving action seems to be something we are morally responsible for as opposed to missing a golf putt, as missing the putt is not an intentional and voluntary action. However, I find that this dichotomy between SFAs and Austin-Style examples is insufficient for showing how we can be morally responsible for SFAs. I say this because even if an SFA was intentional and it was indeterministic which option the person chose during an SFA, it is not obvious that the person is morally responsible. This is because we still do not know what made the agent settle on a particular option over a conflicting one.

During an SFA, an agent might be torn between a morally praiseworthy and morally blameworthy action, both of which are consistent with the agent's past. In Kane's terms, the person could potentially endorse either option in this situation, as both options are consistent with the agent's past. However, if there is no ostensive cause for why one action is chosen over the other, it still seems a matter of luck whether the agent commits a morally praiseworthy or morally blameworthy action. This issue is illuminated by Gregg Caruso when he raises the concern that the event-causal libertarian agent does not have the power to settle whether a decision will occur, and from this, it follows that they lack the desert required for moral responsibility (Carusso, 2015, p25). If Kane is correct in that when the conditions for plural voluntary control are met, the agent has UR and endorses one choice over another, Kane nevertheless cannot account for why one option is chosen over another. Therefore, even if SFAs hold advantages over Austin-Style examples in that they are "freer" and less chancy in virtue of not being involuntary or unintentional, the agent setting their will right there and then does not account for what made the agent settle on one particular option over another right then and there.

A tangential issue arises when Kane attempts to account for how indeterminacy occurs due to parallel processing (Kane, 2002, p419-421, 2007, p28-31). To reiterate, he provided an analogue between solving a mathematical puzzle and solving an indeterminate decision, as in both cases, there is a solving of indeterminate chaos in our brains. I find this analogue overly ambitious. The analogue becomes overly ambitious when Kane claims that similarly to an agent being responsible for solving the mathematical puzzle, the agent is also responsible for the result of their deliberation during an SFA regardless of which option they end up choosing (Kane, 2007, p30). This raises questions, as it is not clear how failing to do the mathematical puzzle would be similar to an SFA in certain regards.

To illustrate this issue, consider the following example. A person is torn between infidelity and faithfulness. In choosing either option, the person has succeeded in doing something they attempted to do as they tried to do both options simultaneously. However, in the case of a mathematical puzzle, I do not believe a person would be trying to both succeed and fail at the puzzle in the same manner. I do not think Kane believes we are attempting to succeed and fail at the puzzle either, though if this is what he claims, it would be glaringly problematic. Instead, I interpret Kane to say that the effort to solve the mathematical puzzle is an analogue to the effort

we exert toward one choice we are trying to make during an SFA. In both examples (Trying to stay faithful and trying to solve the mathematical puzzle), we try to make sense of the world through the chaos in our brains. In both cases, there is noise and indeterminism, either through the competing efforts in making a choice or through our confusion about how to solve the puzzle. This seems a more plausible reading of Kane to me. However, the analogue still falls apart when considering how the failure to complete the puzzle relates to the competing effort in the brain during an SFA, where an agent ends up acting faithfully or unfaithfully. As Kane says, during an SFA the failure to do one thing is not just a failure, but a voluntary succeeding in doing the other (Kane, 2007, p30). I find it implausible that we can say the same thing about failing a puzzle, as we presumably fail at the puzzle due to a lack of ability and it therefore seems unintuitive to call this failure voluntary in any way.

If the agent is not trying to achieve both outcomes in both scenarios, drawing an analogue between the two also seems strange. Especially so when the claim that we are trying to do two things at once during an SFA already raises questions of its own, and because parallel processing is not uncontroversially an incompatibilist idea as others have argued it to be compatible with determinism (Vargas, 2007, p135). Kane has claimed it makes sense to do two competing things at once and that parallel processing not only makes this possible but that this capacity is essential for exercising a free will (Kane, 2005, p138). However, Randolph Clarke has claimed the opposite in arguing that parallel processing is not required for rational actions and, in turn, is not a necessary requirement for free actions (Clarke, 2003, p92). Another issue is that of whether the "stirring up of chaos" in our brains that Kane takes as a basis for his indeterministic view can truly be said to be indetermined or determined, as this is a point that has been disputed both in the scientific and philosophical literature (Vargas, 2007, p143, Kane 2007, p182). Notably, Daniel D. Carr has also expanded on this on Chaos by noting how the quantum events that lie at the heart of Kane's indeterminacy are just as possible in a deterministic world as an indeterministic one (Carr, 2019, p99).

To summarize my critique of this aspect of Kane's account, his analogue is based on us attempting to do two contradictory actions at once. This claim can be contested in a multitude of ways, as it is unclear whether attempting to do two contradictory actions at once is possible, indeterministic, or required for us having free will. It is also not clear whether comparing a SFA

to the solving of a puzzle is a fitting analogue, as it is not clear how we can be said to try to fail at solving a puzzle, even if the outcome it is indeterministic. Furthermore, the basis for this dual effort is based on empirical science that is controversial, as the concept of parallel processing has been interpreted as indeterministic and deterministic by different thinkers, and Kane's view of chaos is based on a view that is not universally agreed upon.

I also find Kane's description of the leadup to SFAs to bring with it a similar type of mysticism that he attempts to avoid in rejecting agent-causal libertarianism. I say this because he assigns a special role to choices when he describes them as "the formation of an intention", and that writes that it is the choice that resolves uncertainty about how we should act next (Kane, 2007, p33). He elaborates on this point by writing the following:

> A choice is the agent's when it is produced intentionally by efforts, by deliberation and by reasons that are part of this self-defining motivational system and when, in addition, the agent endorses the new intention or purpose created by the choice into that motivational system as a further purpose to guide future practical reasoning and action. (Kane, 2007, p34)

This quote encapsulates a central concern of mine with Kane's view. Namely, I find his idea of the relation between choice, intention, and effort problematic.

If intentions are formed by making choices, and this intentionality requires effort, this raises the question of how an intention can be formed as a result of effort. I say this because it is difficult to see how we would know what intention to divert our effort towards if we are yet to form this intention. In response to Kane, this issue has also been raised by Leigh Vicens, who has questioned what it would mean to make an effort to intend to do something (Vicens, 2015, p97). If I sip my coffee, this action would have been preceded by a prior choice and an intention to do so. But any effort I exceed seems to be related to the sipping of the coffee or possibly related to the deliberation of whether I should do so or not. It seems strange then that the formation of the intention to sip the coffee can be said to require effort. I wanted coffee, and therefore I formed the intention to sip my coffee. It does not feel like I exceeded any effort in forming the intention; it did not even feel like forming this intention was something I tried to do. Therefore, I find Kane's view on the formation of intentions questionable, and as his vie of intentions plays a crucial role in his view as a whole, I argue that this also weakens his view as a whole.

In total, I find that together the abovementioned problems holistically shed light on a bigger issue. Namely, a particular and strict set of conditions needs to be fulfilled if we are to have a libertarian sort of free will, which is a concern that Manuel Vargas also has raised (Vargas, 2007, p141). In Kane's case, there would need to be plural voluntary control, we would need to exceed effort to intend something, and there would need to be indeterministic chaos in the brain that works in a specific way. As all of these criteria pose challenges of their own, and Kane's view is founded on his ideas building on one another to form his account of free will, I deem his stance fragile and severely prone to Ockham's Razor.

3.4 The Issues of Endorsement, Ultimate Responsibility & Proto Self-Forming Actions

In this section, I will continue critiquing Kane's account and direct my attention to his view on endorsement, ultimate responsibility, and our first self-forming actions. As pointed out in the previous section, the issue of indeterminism having to occur in a specific manner poses a multitude of problems. The approach in section 3.3 was to oppose Kane's view by pointing out how the foundation for his view is not as strong as it might appear by shedding light on how problematic the various requirements for his event-causal libertarianism truly are. In this section I continue using the same approach, but I will also address some questions by granting Kane's requirements and critiquing how his view remains problematic even when these requirements are granted.

The most telling issue of Kane's view is his use of the term "endorsement" and how his use of the word differs significantly from how we tend to use it in everyday language. Kane believes that when SFAs happen and the conditions of plural voluntary control are met, an agent "endorses" whichever action or choice they end up making (Kane, 2002, p419, 421, 424-425, 428, 2007 p29, 31, 33). My issue with his use of this terminology is that oftentimes, we have an idea of which option we endorse or would endorse over another. This point can be illustrated by again considering the example of being torn between faithfulness and infidelity. Another fitting example for present purposes is that of sticking to or breaking a diet.

In such cases, people find themselves torn between two competing options and struggling to commit to either, but all the while fully endorsing one of the options, labelling it the option they

"should" choose. The pulling power of the option that we "should not choose"; being unfaithful or breaking a diet, can stem from curiosity, ease of access, instant gratification, or hedonism. We have reasons for wanting to choose these options, but this does not mean that we necessarily endorse these reasons. However, while it seems fair to say that most people would not endorse options they "should not choose", people often end up choosing them all the same. When people fall for a temptation and choose options in this manner, I find it beyond strange to say that they endorse this option. Instead, a common response to such an outcome is to say, "I have such weak willpower". In such cases, I claim we opt for what we find to be the incorrect choice, voluntarily and intentionally, but not because we endorse this option. Therefore, even in granting Kane his claim that a choice of this kind is voluntary either way (Kane, 2002, p412), it does not follow that the choice is endorsed either way. If he believes we endorse all of our choices in an everyday manner of the word, this seems outright mistaken in light of the abovementioned examples. If he is attempting to say something else, I argue that his account of endorsement is in severe need of revision.

As Kane's account of endorsement is one of the requirements he sets for ultimate responsibility, this brings into question whether we can be said to have ultimate responsibility at all. Especially since the idea that we can be said to be ultimately responsible has been questioned by multiple notable philosophers. Gregg Caruso outlines how the incoherence of ultimate moral responsibility was first introduced by Nietzsche, and then later fleshed out by Galen Strawson in the contemporary free will debate (Caruso, 2018, 2.3). These skeptical claims are based on the idea that for us to be ultimately responsible, we also have to be the causes of ourselves, which in turn can be taken to be impossible. Strawson expands on this claim by presenting what he calls the basic argument (Strawson, 1994, p5-7). A simplified version of this argument can be laid out as follows:

> We act in a certain way because of the way we are, and in order to be ultimately responsible for our actions we need to be ultimately responsible for the way we are. However, we are not responsible for the way we are, and thus not ultimately morally responsible for our actions. (Carusso, 2018, 2.3, my wording)

I find the basic argument robust. Even if one holds that we are ultimately responsible for our actions, it is hard to pinpoint where the argument goes wrong. The argument's soundness seems

indisputable, which leaves us the option of attacking the premises if the argument is to be disputed.

Considering that self-formation is a core aspect of Kane's libertarian view, it is no surprise that he finds the basic argument invalid. Kane concedes that the premises of the argument seem sound. However, he also claims that even though we are not the original creators of our characters, we can nevertheless freely change our characters as we mature, which seems to him like a piece of common sense (Kane, 2005, p73). I would agree with Kane on this point if it were not for the fact that he claims we change our characters when we mature "freely". The fact that we change our characters with age seems like a piece of common sense to me as well. However, the claim that we do so freely seems to beg the question when he argues that we have a libertarian form of free will. If the way in which we change our character is determined, Kane and I would agree as incompatibilists that we cannot change our character freely. Yet, Kane himself says that we cannot know if determinism holds true, and that we might never find out if determinism holds true (Kane, 2007, p181). Due to this uncertainty concerning the truth of determinism and the fact that incompatibilism must hold true for us to change our character freely in the manner Kane describes, I deem Kane's claim that it is common sense that we can "freely" change our character with age problematic.

The abovementioned issue of whether we can change our character freely leads us to the question of whether we can begin to form our character freely. To reconcile this question with Kane's account, beginning to form our character freely requires an account of how we can be held responsible for our first SFA and how this action can be an SFA to begin with. After all, the disposition we have formed through previous SFAs makes us ultimately and morally responsible for both further SFAs and non-SFAs. This raises the question of how we can begin making self-forming actions, as they would always require a previous SFA.[26] Pereboom raises this problem of first SFAs the following way:

> ..consider the first free choice an agent ever makes. By the above argument, he cannot be responsible for it. But then he cannot be responsible for the second choice either, whether

---

[26] For another perspective on why this is an issue, see Derk Pereboom's argument on how first self-forming actions do not lead to moral responsibility regardless of whether determinism or indeterminism holds true (Pereboom, 2001, p49).

or not the first choice was character-forming. If the first choice was not character-forming, then the character that explains the effort of will for the second choice is not produced by his free choice, and then by the above argument, he cannot be morally responsible for it. Suppose, alternatively, that the first choice was character-forming. Because the agent cannot be responsible for the first choice, he also cannot be responsible for the resulting character formation. But then, by the above argument, he cannot be responsible for the second choice either. Since this type of reasoning can be repeated for all subsequent choices, Kane's agent can never be morally responsible for effort of will. (Pereboom, 2007, p109)

In answer to this counterargument, Kane says that early SFAs are still possible. This is because even if there is a lesser degree of responsibility present in these SFAs, they are nevertheless self-forming due to the effort exceeded and the endorsement of the outcome.

..responsibility accumulates in human beings as they get older and build up a backlog of self-formed character. The only exceptions to this twofold source of responsibility are the earliest SFAs of childhood in which a backlog of self-formed character does not exist. In these earliest SFAs, all the responsibility is thus in the effort itself and the endorsement of its outcome by the agent when it succeeds. But precisely for this reason the responsibility for these earliest SFAs of childhood is not as great as later ones; in fact, in the earliest SFAs responsibility is minimal. That is why we hold very young children to be far less responsible than older ones and adults. (Kane, 2007, p174-175)

This way, Kane explains why we hold people more accountable with age, an idea for which he has garnered support (Lemos, 2015, p144). However, from my critiques of Kane's view on effort and endorsement, I do not find this to be a satisfactory response to the bigger issue of first SFAs. As mentioned above, Kane concedes that the responsibility of early SFAs is not as great as later ones. Given that I find Kane's account of how we can assign responsibility to an agent due to SFAs problematic, it thus becomes all the more difficult to see why early SFAs should invoke any degree of responsibility at all. Murday illuminates this issue further by pointing out the possibility that the responsibility might instead lie in the agent's character, rather than in the agent herself.

Kane's account is thus that the efforts of will leading up to a later SFA should typically have two features: (a) they constitute part of a character for which the agent was already responsible in virtue of having formed it in some earlier SFA, and (b) the effort is authenticated in the later SFA when it succeeds, in virtue of the agent successfully making the choice to act on that effort and not regarding that choice as a mistake or accident.

Kane holds that when the first self-forming action satisfies (b), partial responsibility is conferred on the agent in virtue of her endorsing the outcome. But why is that endorsement attributable to the agent, as opposed to being attributed to the character for which she is not responsible? If the endorsement itself is not attributable to the agent, then it may be hard to see why it confers any responsibility on the agent (Murday, 2017, p1322, my formatting)

In other words, if the responsibility for the action can be assigned to a person's character, and the person is not responsible for this character, it is difficult to see how the person is responsible. This reverts us to the age-old question raised by Aristotle of how we need to be responsible for having formed our disposition if we are to be held responsible for the actions we perform because of our disposition. I, therefore, argue that as Kane's idea of endorsement and effort is problematic, and as effort and the endorsement of an action is required to ever start performing SFAs, his entire view on how we can be said to have free will and be morally responsible is weakened further. Combined with the other issues raised in these sections, I find Kane's view holistically implausible, and I, therefore, find his event-causal libertarian account unconvincing.

4. Hard Incompatibilism

4.1 Hard Determinism & Hard Incompatibilism

I have now shown why I find both libertarian and compatibilist answers to the question of whether we have free will and moral responsibility to be unconvincing. As these are the most common pro-free will accounts, I am skeptical of whether we can be said to have free will and whether we can be held moral responsibility. I will proceed by discussing what type of position regarding the question of whether we have free will and moral responsibility is worth endorsing with this skepticism in mind. One possible answer is to endorse hard determinism. Hard determinism denies that we have free will and affirms determinism, as opposed to soft determinists who are compatibilists that believe that the forms of free will worth having are not undermined by determinism (Kane, 2005, p69). At first glance, hard determinism can seem like the position we are inevitably left to endorse if we are to take a skeptical stance in the free will debate, as historically incompatibilists have either been libertarians or hard determinists (Vihvelin, 2022, 1). However, modern hard determinists are few and far between (Kane, Fischer, Pereboom, Vargas, 2007, p3). Instead, Carusso regards most positions defended by skeptics today as successors to traditional hard determinism and makes the claim that the reason the position has few defenders today is that the standard interpretation of quantum mechanics casts doubts on universal determinism (Carusso, 2018, 2.1).

Quantum states, meaning states in which a quantum mechanical system can be, assign definite values to all physical quantities, and in the standard formulation of the theory, probabilities are built into the system (Myrvold, 2022, 2.1).[27] In other words, it seems, at least on the surface of it, to be observable indeterminacy in nature. Examples of this type of indeterminacy in nature include radioactive decay and photon emission, which, according to many physicists, can only be given a probabilistic description (Hoefer, 2023, 4.4). If it were true that these phenomena could only be given a probabilistic description, it would seem to follow that they cannot be said to be fully deterministic either. If this is the case, I would be hard-pressed to endorse a hard determinist view on the question of free will. However, the claim that indeterminacy follows

---

[27] Quantum indeterminacy and quantum randomness have been widely discussed, such as in the works of Falkenburg & Weinert, Paterek & Collogues and Jager (Falkenburg, Weinert, 2009, Paterek & Collogues, 2011, Jager, 2017).

from what we know about quantum states is inconclusive, and some even claim the opposite.[28] In other words, science has not uncontroversially proven indeterminism or determinism to hold true. Nevertheless, whether it be due to solid reasons for believing there is indeterminacy in nature or due to the difficulty of proving this question either way, a hard determinist view has become less appealing as a result.

The result of the abovementioned shift towards the view that determinism might very well be false has led contemporary skeptics to generally favor a hard incompatibilist stance instead of a hard determinist one (Talbert, 2019, 3.3). Just like libertarians, hard incompatibilists maintain that free will and determinism are incompatible, and the stance is "hard" in the same sense as hard determinism is, as proponents of hard incompatibilism claim that true free will does not exist (Kane, 2005, p71). Therefore, defenders of the stance can be skeptical about us having free will while simultaneously being agnostic about whether determinism holds true, which Derk Pereboom is (Pereboom, 2007, p85). Agnosticism concerning determinism, while also endorsing incompatibilism and free will skepticism, also brings with it an answer to a key question in the free will debate. Namely, how can we be said to have free will or be morally responsible if the world is indeterministic? A hard incompatibilist will answer by claiming that we are not morally responsible or have free will and this view is therefore not prone to the luck objection.

When discussing compatibilism in chapter one, I showcased how detrimental determinism was to our notion of free will. However, as I discussed when addressing the luck objection in chapters 2 and 3, our actions could be a matter of chance, even if they were indeterministic, and, therefore, we would not be morally responsible for them. Hard determinists will fundamentally disagree with libertarians about whether determinism holds true. This is an impasse that is difficult to get past, as in citing John Earman, Kane points out how we are a long way from knowing if determinism holds true and that we may never definitely know (Kane, 2007, p181). Hard incompatibilists avoid this issue, as they can acknowledge the possibility of indeterminism holding true while maintaining that we are not free or have moral responsibility if indeterminism holds true.

---

[28] Hoefer writes that the idea that the world is ultimately chancy at the microscopic level is partly mistaken, misleading, or both. He also points out that whether quantum mechanics can be taken to support an indeterministic or deterministic worldview is dependent on what interpretational or philosophical decision we adopt (Hoefer, 2023, 4.4).

Pereboom expresses this point by writing that if determinism holds true, we would not be morally responsible and that if indeterminism were true and events and states were the only causes of our actions, we would also not be morally responsible for our actions (Pereboom, 2007, p85).[29] Pereboom, when specifying that the causes for morally responsible actions cannot be events and states, rules out that we can be held morally responsible in the way event-causal libertarians claim. This sort of hard incompatibilism does not rule out the possibility of us being morally responsible in some cases. But, if we are to be held morally responsible at all, it would have to be for distinct reasons from those endorsed by event-causal libertarians like Kane. As discussed in Chapter 2, Pereboom instead finds that we would need agent-causal powers to be held morally responsible, but as it is unlikely that we have agent-causal powers, there is also little reason to believe we are morally responsible. I agree with this point as I do not believe event-causal libertarianism can be endorsed for the reasons outlined in Chapter 3. I also agree that if agent-causation could somehow be proven to hold true, we would have good reasons to believe we have free will. Therefore, while new theories on free will might get developed, I have difficulties seeing how anything but us having agent-causal powers could serve as a sufficient explanation for us having free will, at least when considering the current discourse on free will and moral responsibility.

So far in this thesis, I have discussed compatibilism, finding it both insufficient at providing an answer to the question of how we can have free will and moral responsibility and unconvincing in providing reasons to abandon incompatibilism. I have considered libertarianism as an alternative, and while I agree with this view on the incompatibility of free will and determinism, it did not provide a sufficient answer to how we can be held morally responsible or have free will. Hard determinism accounts for my skepticism concerning free will while retaining an incompatibilist stance, but this view might, for all we know, not be reconcilable with the laws of nature and creates an intelligibility question of its own. For these reasons, I endorse hard incompatibilism, which accounts for the abovementioned issues and because I find it the most plausible and unproblematic view in the free will debate in virtue of how unconvincing I find the other commonly endorsed options. The rest of this thesis will be dedicated to explaining my

---

[29] Pereboom is here referring to moral responsibility of a "basic desert" kind. This means that an agent who performs an action that belongs to her in such a way that she deserves praise or blame for it is morally responsible for the action if she understands the action's moral status (Pereboom, 2007, p86). He calls this desert "basic" because the agent deserves praise or blame solely in virtue of this fact.

reasoning for why I endorse this view further and to explain how skepticism concerning moral responsibility and free will does not lead to the type of negative consequences that many philosophers have claimed. On the contrary, I will argue that an unjustified belief in free will and moral responsibility can have determinantal consequences of its own. I will also argue that believing in hard incompatibilism is not an undesirable stance to adopt on the question of whether we have free will and that hard incompatibilism holding true would not be an undesirable outcome.

4.2 Decisions & Uncertainty

Even if one were to concede that libertarianism and compatibilism poses issues of their own, it can be argued that hard incompatibilism poses even bigger ones. Most notably, we obviously seem to have free will and we seem free to do what we like when we deliberate and decide. Our fundamental worldview can be said to be shaped by the idea that we are free, with van Inwagen writing that if some unimpeachable source told him he did not have free will, he would regard it as proof he did not understand the world at all and that this would be to accept a mystery (van Inwagen, 2015, p283). I agree this is a mystery, though van Inwagen and I also agree that accounting for how we have free will is a mystery of its own. Van Inwagen goes on to write that different people might judge the size of these mysteries differently, with him claiming that he accepts the belief in free will as the lesser mystery (ibid, p284). However, I claim the opposite for just the same reason. In this section, I will outline why van Inwagen's reasoning for calling a lack of free will the greater mystery is unconvincing and how our ability to deliberate and decide is not threatened by hard incompatibilism to any meaningful extent.

Conflicting viewpoints on which of these two mysteries is bigger can be in danger of remaining at an impasse. However, van Inwagen does well in expressing what lies at the heart of his inclination to gauge a lack of free will as the greater mystery. To make his point, van Inwagen uses a thought experiment first laid out by John Locke:[30]

---

[30] With this argument, van Inwagen is defending deliberation incompatibilism. This view holds that there is an incompatibility between agents rationally deliberating between performing mutually exclusive actions and agents believing that there are prior conditions that render the performance of either one of these actions impossible (Henden, 2010, abstract).

Suppose, for example, that you are in a certain room, a room with a single door, and that this door is the only possible way out of the room. Suppose that, as you are thinking about whether to leave the room, you hear a click that may or may not have been the sound of the door's being locked. You are now in a state of uncertainty about whether the door is locked and are therefore in a state of uncertainty about whether it is possible for you to leave the room. Can you continue to try to decide whether to leave the room? It would seem not. (van Inwagen, 2015, p282)

This thought experiment concerns a situation where you are unsure if you have the ability to do something. Similarly, I am unsure if I can type 60 words a minute; I will likely be a little too slow or barely fast enough. Therefore, just like I can't decide to leave the room, I cannot decide to write 60 words a minute. So, van Inwagen seems to be correct in saying that it does not seem like I am able to decide whether to leave the room, as I don't know if I could do so. The thought experiment is appealing to van Inwagen because it demonstrates how uncertainty about the ability to perform an action means we cannot decide to do it, and in turn, we need to believe we have free will in some cases. If we didn't, all actions would fall into this category of uncertainty where we could not decide to do anything. The result of this, John Martin Fischer claims, would be that the distinctions we commonly draw in law and common sense would be put in jeopardy of disappearing (Fischer, 2007, p45).

Van Inwagen goes on to explain how a lack of belief in free will is not something we can do without and why his conviction of this follows from Locke's thought experiment:

This thought-experiment convinces me that I cannot try to decide whether to do A or B unless I believe that doing A and doing B are both possible for me. And therefore I am convinced that I could not try to decide what to do unless I believed that more than one course of action was sometimes open to me. And if I never tried to decide what to do, if I never deliberated, I should not be a very effective human being. In the state of nature, I should no doubt starve. In a civilized society, I should probably have to be

institutionalized. Belief in one's own free will is therefore something we can hardly do without (van Inwagen, 2015, p282).[31]

Van Inwagen claims that it is impossible to decide whether to do A or B unless he believes "both are possible to him". In one way, this statement seems plausible. I can decide whether to sip my coffee as there is a presumption that I would succeed in either doing so or abstaining from doing so. I cannot decide whether I should change the laws of nature, as this is impossible. I cannot decide to leave the locked room, not due to this being impossible, but because of the uncertainty of whether I can do so or not. I can decide between possible options but not between impossible or uncertain ones. It seems then that if all choices were uncertain in this manner, no one could decide between anything as we would not know what is possible.

However, I believe van Inwagen is mistaken on multiple accounts in his interpretation. The reason I say this is because if determinism holds true, the implications of Locke's thought experiment would be different than if we did have free will, and I claim the implications of this have been overlooked. Furthermore, I believe van Inwagen's claims about how we would respond if we genuinely believed we did not have free will are unrealistic. I will proceed by laying out an argument to showcase how van Inwagen's interpretation of the thought experiment stipulates free will to make a point about how things would be if we did not believe we had free will and why his interpretation, therefore, begs the question.

I believe there is a crucial difference in two forms of uncertainty that van Inwagen is not properly accounting for in his reasoning:

> U1 – Uncertainty about whether one would be capable of doing A or uncertainty about whether one would be capable of doing B (In some situation at some time).

> U2 - Uncertainty about whether A is possible for one to do or uncertainty about whether B is possible for one to do (In some situation at some time).

If we have free will, we have both U1 and U2 sometimes. Sometimes A or B would be deemed possible, and sometimes A or B would be deemed impossible. In neither of these cases would we

---

[31] A similar sentiment has also been expressed by Daniel Dennet, who writes that a lack of belief in one having free will would result in an attitude towards actions and choices which would disable one as a chooser (Dennet, 1984, p184).

have U1 nor U2. This is because if we deem doing A or B possible or impossible, we would not be uncertain of whether we could do A or B. In other cases, we would be uncertain about whether we could do A or B, and we would have both U1 and U2 in these cases. This is because we would doubt whether doing A or B was possible due to being unsure whether we could do A or do B, and we would doubt whether we could do A or B because we are unsure whether doing A or B is possible. It makes sense then that van Inwagen would equate U1 and U2, as from a libertarian standpoint, U1 would follow from U2, and U2 would follow from U1. In Locke's room, the person in the room has both U1 and U2 as they are uncertain about their ability to exit the room and, therefore, do not know if it is possible for them to exit the room.

However, if we do not have free will, U1 does not follow from U2 in this manner. If we do not have free will, we will always have U2, but not always U1. We will always have U2, as we cannot know what is possible because we do not know whether we will choose to do A or choose to do B. Everything might be determined and out of our control, or some indeterminacy we have no agency over might cause us to do A or do B. Either way, we cannot know what is possible as we can only do either A or B, and the option we do not choose will, by necessity, be impossible. It is impossible because this option requires us to will something we would not have willed, as I argued was the case if determinism holds true in chapter 1.4. However, we would have U1 sometimes in the same manner as if we did have free will. For example, I would not be uncertain about whether I would have been capable of sipping my coffee. If I had tried, I would have succeeded if that had been what I had chosen to do, and I would not have U1. However, I would be uncertain whether I could type 60 words a minute, and I could not decide to type 60 words a minute whether this decision was free or not. Therefore, I would have U1. In cases of uncertainty about ability, both U1 and U2 would still exist without free will, such as in Locke's room. However, given that we would always have U2 if we did not have free will, but we could still lack U1 if we did not have free will, U1 would not follow from U2 if we did not have free will.

I believe the fact that there are U1 and U2 in Locke's room even if there is no free will, combined with the fact that U1 and U2 follow from one another on a libertarian view, has led van Inwagen to overlook this relation between U1 and U2. Locke's thought experiment seems potent because in it, there are U1 and U2 regardless of whether we have free will, but the real-world consequences van Inwagen derives from the experiment are based on the incorrect

assumption that there would always be U1 and U2 if we did not believe we had free will. Consider van Inwagen's claim that we cannot decide whether to do A or B without believing both options are possible for us (van Inwagen, 2015, p282). Suppose I lacked free and did not believe I had free will. If I decided to sip my coffee, I would not be incapable of making this decision despite being uncertain about whether this was possible in virtue of determinism or indeterminism. I did have U2, as I cannot know what I will end up willing and I cannot do what I do not will, but I was nevertheless able to deliberate. However, I did not have U1 as I believed I would succeed in sipping my coffee If that is what I ended up willing to do. Therefore, despite me being uncertain about whether sipping the coffee is possible, I can still deliberate between the options and make a decision as I did not have U1, even though that decision might not be freely willed and I had U2. I see no good reason that being made aware of this lack of free will would change anything about scenarios of this kind. Therefore, the claim that such an inability to "deliberate" would result from U2 and a lack of belief in occasional alternative options seems unjustified. The same can be said of van Inwagen's claim that it would follow that I would probably be institutionalized if I always had U2 (van Inwagen, 2015, p282).

Above, I have argued for where I believe van Inwagen's interpretation has gone wrong. However, the core problem of his view can also be laid out in a more simplistic manner. Ask yourself, if you stopped believing that A and B are both possible to you or you stopped believing that alternative possibilities are sometimes available, would this result in you being unable to deliberate? Would this uncertainty make it impossible for you to ever decide on one option over another? Would every single "choice" you encountered bring with it the same type of uncertainty as if you were locked in Locke's room? Would this, in turn, lead to behavior that would likely have you institutionalized? Personally, I cannot fathom how it could. To illustrate why, consider what I view as a strange and unjustified yet common response to the hypothetical scenario of finding out one does not have free will. This response is to give up on life, not seeing the point of continuing without free will. Internalizing this decision, one might decide to lie in bed and decide against going to work anymore, not seeing the point of doing so. Taking this point to the extreme, one might even contemplate whether it is worthwhile to continue living. If I were to assume, based on van Inwagen's claim that he would not feel like he understood the world at all if he was made aware that he lacked free will, he would presumably deliberate over his worldview. However, in all these examples, the ability to deliberate has not been lost. All the

above responses to a lack of belief in free will hinge on deliberation and decisions of some kind. Therefore, I argue, this deliberation is possible as a hard incompatiblist and any consequences that would result from a lack of belief in free will could not possibly be of the kind van Inwagen describes.

4.3 Praiseworthiness, Gratitude & Love

In the previous section, I discussed the inclination that we have free will, and I argued that a lack of such a belief would not make us unable to deliberate or lead to severe consequences of the kind van Inwagen describes. I will proceed by examining how a lack of free will and moral responsibility would impact our interpersonal relationships and how some philosophers have argued that these aspects of our lives are threatened by skeptical views. I will also discuss how we could not justifiably praise or blame each other on a hard incompatibilist view and argue that this is far from as detrimental of an outcome as others have claimed. Then, I will discuss love and gratitude as examples of critical interpersonal attitudes and how these concepts can largely be maintained on a hard incompatibilist view.

If you are like most philosophers, chances are you find the idea of free will skepticism unattractive.[32] Kane calls the skeptical positions of hard incompatibilism and hard determinism rails that most people do not want to touch in fear of being electrocuted (Kane, 2005, p71). A key reason why people have this view on skeptical positions on the question of whether we have free will and can be said to be morally responsible is that skeptical positions appear to threaten important aspects of our everyday lives. For example, Kane points out that it can be questioned whether we can value love without having free will (Kane, 2005, p76-77). Another example is Peter Strawson, who writes about how our ingrained reactive attitudes, such as gratitude and resentment, are dependent on the act that evoked the feeling being blameworthy or praiseworthy (Strawson, 1962). This, in turn, requires the person who evoked the attitude to be morally responsible. If they are not, the lack of these attitudes would threaten some of the foundational notions that shape our interpersonal relationships.

---

[32] Upon being surveyed in 2020, only 11.2% of academic philosophers accepted or leaned towards believing that we do not have free will (Chalmers & Bourget, 2023).

William Jaworski points out how our idea of moral accountability, praise and blame is based on the assumption that we have free will and how we take this for granted in everyday attitudes (Jaworski, 2011, 13.3). Saul Smilansky, despite finding free will illusionary, still reminds us that our social interactions and emotional lives would be wounded if we realized and internalized the idea that our assumption that we have libertarian free will is false (Smilansky, 2000, p277). Lastly, Pereboom raises the issue of how a lack of free will could also be used as an excuse for poor behavior. He points to how he believes Sartre would condemn such behavior as we would be acting in bad faith if we denied our freedom.

> However, what should we say to someone who regularly and deliberately does wrong, refuses to make a commitment to morality, and offers hard incompatibilism and his consequent lack of freedom as an excuse for his behavior? Wouldn't the hard incompatibilist have little to say to such a person? Jean-Paul Sartre would impugn such denial of freedom as a form of "bad faith," a kind of self-deception, and he is clearly describing a form of thought and behavior that we would want to avoid. (Pereboom, 2001, p155)

However, there are multiple answers to these claims about free will skepticism that result in a significantly less gloom of an outcome than the abovementioned thinkers claim. These answers mainly concern how we can still retain many aspects that we deem key for our interpersonal relationships and self-worth without free will and moral responsibility.

Some of these answers will require a revisionist approach to the concept of moral responsibility. Revisionism is a stance of its own in the free will debate, and in simple terms, the view advocates that what we tend to think about free will and moral responsibility differs from what we ought to believe about these two concepts (Vargas, 2007, p127). In other words, our concepts of free will and moral responsibility need revision; we should reconsider how we think about them. One can also endorse any one position in the free will debate while simultaneously taking a revisionist approach. According to himself, Pereboom's version of hard incompatibilism is highly revisionist regarding the notion of moral responsibility (Pereboom, 2007, p122). Therefore, some of the arguments he lays out in favour of this position that I will defend requires us to reconsider how we think and ought to think about the notion of moral responsibility to begin with.

I will first examine the implications of claiming that there is no moral responsibility and that, as a result, we cannot justifiably morally praise or blame people for their actions. Consider the issue raised by Pereboom above of how Sartre would impugn a person's refusal to make a commitment to morality and justifying such an attitude on the grounds that one is a hard incompatibilist. On the surface, this lack of commitment to morality seems to stem from a lack of guilt. Pereboom writes that guilt is dependent on a belief that one is blameworthy to begin with and that guilt, therefore, also appears to be a prerequisite to motivate repentance (Pereboom, 2007, p120).[33] This raises problems for hard incompatibilism. Namely, given that we could not justifiably be morally blamed as hard incompatibilists, why would we then ever feel guilty? Furthermore, if we do not feel guilty, how could we be motivated to repent for our wrongdoings? An answer to the latter question can be given by conceding that feeling guilty might not make sense if we lack the type of freedom required for moral responsibility but that moral improvement is possible without the notion of guilt. This is because there are other central emotions and attitudes we can rely on for moral improvement following wrongdoing. Pulling from Bruce Waller, Pereboom writes how hard incompatibilism is compatible with adopting the following type of attitude to solve this problem:

> you acknowledge that you have done wrong, you feel sad that you were the agent of wrongdoing, and you deeply regret what you have done (Waller 1990). Also, because you are committed to doing what is right and to your own moral improvement, you resolve to refrain from behavior of this kind in the future, and seek the help of others in sustaining your resolve. (Pereboom, 2007, p120)

I do have one issue with this claim, as regretting what you have done seems to me redundant if we were not free to do anything but what we did. But aside from the point made about regret, I otherwise agree with this sentiment as it shows how we can attain moral improvement through acknowledgement, sadness and help as hard incompatibilists.

I believe we apply this type of attitude in many everyday settings where we do not consider the situation to be of moral severity. For example, let us imagine that I accidentally broke my coffee

---

[33] If we accept praise and blame as central notions that drive our moral improvement, it seems to me that giving up either would be different sides of the same coin. However, van Inwagen refers to the loss of blame as attractive and the loss of praise as unattractive (van Inwagen, 2015, p283), which I find misses the mark on this issue.

machine. I then acknowledge that this was a mistake, so I refrain from doing such an action that caused it to break in the future. If I don't understand why I caused it to break, I might consult others to find out why I broke it to avoid breaking it again. It is up to me to improve the situation by fixing my coffeemaker and not breaking it in the future, or neglecting to do anything and thereby failing to improve the situation. All of this is possible without feeling guilty about breaking the coffee maker in any way. In more direct wrongdoings, especially those concerning interpersonal relationships, we might feel a deeper sadness and find it more personal to be the agent of wrongdoing. However, the same type of attitude as in the case of the coffeemaker of not being guilty but still showing moral improvement can clearly be applied to wrongdoing in general. Therefore, I agree with Pereboom that hard incompatibilism does not challenge moral improvement despite guilt and blame being unjustified on this view.

A more nuanced issue was raised by Smilansky above, in that a lack of belief in libertarian free will would wound our social interactions and emotional lives. Kane and Strawson also provided love and gratitude as examples of notions that seem to rely on the idea that the person who invoked the attitude acted freely. I will proceed by considering how the notion of gratitude is not threatened by hard incompatibilism in the manner the above thinkers claim. The way many people might think about gratitude under the assumption that we have free will and are morally responsible is best exemplified by someone receiving a gift they did not want but are nevertheless genuinely grateful for. Someone thought about them and expressed goodwill towards them, which provides the person with joy and a reason to be grateful. Being appreciative of the goodwill assumes this goodwill was freely expressed; someone willed you good. If there was no goodwill, all you would have is a gift you did not want. In other words, "It is the thought that counts", but the thought counts because we assume it to be free, and in turn, feel gratitude. For this reason, it might seem like we would not be able to retain this sense of gratitude as hard incompatibilists.

However, Pereboom argues that as gratitude involves being thankful towards someone, there are cases where we show this type of thankfulness without also believing that the person was morally responsible for the action (Pereboom, 2007, p121, 2014, p190). Pereboom exemplifies this by noting how we can show the type of thankfulness required for gratitude towards small children for some kindness they did without holding the child morally responsible (ibid). I agree,

and this point can be expanded upon in the following way. If we were to be held morally responsible, compatibilists have argued that we need to be rational (Dennet, 1984), sane (Wolf, 1987), or act without being coerced (Fischer, 2007, p45-48). However, I can imagine an irrational, insane or coerced person committing an act of goodwill towards me, which I can be thankful for while knowing fully well that the person is not in a state where these compatibilists claim the person can be held morally responsible. I don't have to mind that the person showing me goodwill was in an irregular state where most would exempt them from praise or blame. I can still be thankful for the joy their action brought me. An insane person I know to be insane can help me fix my car when it breaks down, and I can drive off thankful that he did this without holding them morally accountable for this, just like I would not hold them morally responsible if they did some injustice towards me due to their insanity. A parent can be thankful that their small child did not throw a tantrum during dinner, but if the child did throw a tantrum instead, the parent would not hold their small child morally responsible for it.

If moral responsibility is not a necessary condition for feeling thankful, and this thankfulness is a major part of gratitude as we know it, we could retain a clear sense of gratitude as hard incompatibilists. I say this because even if actions are never morally praiseworthy due to our lack of free will, I believe the abovementioned examples of feeling thankful could also be applied to general acts of goodwill and kindness. On gratitude, Pereboom also writes that we would retain our sense of joy as a response to beneficent acts, as there is nothing about hard incompatibilism that undermines joy as a response to generous acts on one's behalf (Pereboom, 2007, p121, 2014 p190). This point also seems intuitive, as we might feel joy similarly to how we feel thankfulness in the face of an irrational or compulsive act that resulted in goodwill. I can feel joyful that the insane person helped me fix my car, and the parent can feel joyful that their child did not throw a tantrum. For these reasons, I believe that gratitude could be retained to a large degree on a hard incompatibilist view. We could retain the positive emotions gratitude evokes, and we could still direct gratitude towards others as this is something we already do towards people we do not normally regard as morally responsible. If no one is ever morally responsible, I thus find it unproblematic to claim that we could universally adopt this attitude on gratitude.

An even more significant notion than gratitude that can seem threatened by denying free will and moral responsibility is that of love. William Jaworski points out why by illuminating how some revisionists have expressed a need to revise our concept of love if we lack free will:

> Some revisionists claim that many people will need to revise their expectations about having loving relationships if there is no free will or moral responsibility. People who think that genuine loving relationships must be freely chosen in a libertarian sense, for instance, will have to revise their ideas about whether such relationships are really possible, and in that case, they will also have to revise any behavior that is directed toward achieving such relationships since it would be irrational to pursue a goal that was unattainable. (Jaworski, 2011, 13.29)

On the surface, this need for revision seems plausible. To illustrate why, imagine a case where A and B feel love for one another, and it is the case that we have free will. A freely wills to pursue a loving relationship with B, and B freely wills to pursue a loving relationship with A. The combination of free will, love, and both parties willing to form a loving relationship seemingly makes it possible to develop this loving relationship. Intuitively, if this loving relationship was not formed freely, it can be questioned whether this is an instance of love at all.

However, I consider the abovementioned conditions as sufficient, but not necessary, for having a loving relationship. I say this because we can show love in a multitude of other ways and because I do not believe that A and B choosing to love each other freely, rather than merely deciding to love each other, subtracts anything substantial from their loving relationship. To illuminate the point of us being able to show love in a multitude of ways, Pereboom again uses the example of small children. He argues that children rarely get loved because they exercise a freedom of a certain type or that they deserve love because of their free actions being free, and he goes on to say that when adults love one another, this is rarely the case either (Pereboom, 2001, p202, 2007, p121, 2014, p190).

I believe the point he makes on why we can show love for children is similar to how we can show gratitude towards them. However, his point on why adults might love another in a similar way might strike some as odd. To make this point, he lists several reasons why he believes this to be the case:

1. Morally admirable qualities can be regarded as lovable even if they are not praise or blameworthy.
2. We love each other for a complex range of reasons (Including intelligence, appearance, style, etc.)
3. We sometimes attempt to restore or maintain love while wanting the other party to make a decision to reciprocate, not for the other party to specifically make a decision to reciprocate freely.
4. The lack of justification for attitudes such as resentment and guilt that would be lost without moral responsibility are not required for loving relationships. (Pereboom, 2007, s121-122, my formatting and wording)

The first point seems intuitive. A person might be caring and compassionate, and these characteristics can be deemed morally admirable. However, these characteristics can be deemed admirable, and we might appreciate these qualities and aspire to adopt them ourselves but do so regardless of whether these characteristics were freely developed and are morally praiseworthy. The second point illuminates how other factors are at play in love, aside from whether the love was willed freely. Again, we can gravitate towards loving someone based on characteristics such as intelligence and appearance, even though these factors are not characteristics a person inhabits due to choosing them freely. I find the third point somewhat redundant. While I agree with the point, I think it does little to alleviate the opinion of those who take free will to be needed for love in the first place. I say this because desiring a decision to reciprocate an attempt to restore or maintain love does not seem much different from wanting someone to choose to love them.

The fourth point builds on the idea that attitudes such as guilt and blame would not be justifiable if we lack moral responsibility. To me, this is the most potent point. Not only does it say how loving relationships do not require guilt and blame, but it illuminates the lack of justification for holding these attitudes to begin with. I say this because I believe that if these attitudes were never justified to begin with as we lack moral responsibility, we can develop healthier relationships by discarding them. If we are not justified in assigning blame or guilt but nevertheless assign them in loving relationships, these loving relationships would suffer from unnecessary negative emotions. Not only can giving rise to such attitudes unjustifiably be regarded as inherently bad, but it can be detrimental to the relationship and, in some cases, be a cause of it ending. For these

reasons, I claim that love would not be threatened to a significant degree by a lack of free will, but rather that a belief in free will and moral responsibility can itself be a threat to loving relationships.

4.4 Justice & Real-Life Implications

Not only is the question of whether we have moral responsibility closely tied to our interpersonal relationships, but it also has major implications for whether it is permissible to punish others. Therefore, whether we have moral responsibility is also crucial for how justice systems are conducted. In denying moral responsibility, hard incompatibilists have to account for the impact this would have on punishment and justice. This is because it would be unacceptable to blame criminals for what they have done, and if we cannot blame criminals, this jeopardizes what is arguably some of the core features of morality and justice (Pereboom, 2001, p158). In turn, this could leave us without legitimate methods of preventing people from doing terrible things to one another (ibid, p159).

However, I believe justice and a revised form of punishment are reconcilable with a hard incompatibilist stance. There are tensions between different theories of justification of punishment and practical attitudes on questions concerning what types of punishment are morally acceptable (Tebbit, 2017, p242). I believe hard incompatibilism can illuminate the core issue of why this is the case and provide an answer to how we ought to think about punishment. Therefore, in this section I will argue that justice systems can be revised and fulfil their purpose if we do not have moral responsibility. I will also argue that conducting justice under the presumption that we are morally responsible can have severe consequences of its own if it turns out that we do not have moral responsibility. Most glaringly, the issue stems from how closely tied the ideas of guilt and free will are in terms of how crimes are judged:

> To be held guilty in mind, the defendant has to have known what she was doing, to have done it intentionally and to have carried it out of her own *free will*. This is the norm, and there is a strong presumption in its favour. (Tebbit, 2017, p193, my italics)

Put another way, if we do not have free will, it follows that there would be a strong presumption that people cannot be held guilty in mind either. However, I do not believe that we should

abolish justice systems if we do not have free will or we cannot be held morally responsible. Instead, I argue that the justification for serving punishment should be revised if we do not have free will, as punishments that are carried out on the grounds that criminals acted with free will and are therefore guilty would be unjustified.

Pereboom argues that we can turn to moral admonition and encouragement instead of threatening people as if they deserve blame (Pereboom, 2007, p115). He proceeds to argue that justifying criminal punishment on the grounds that it educates criminals morally is not challenged by hard incompatibilism in particular and that deterrence theories that justify the punishment on the grounds of self-defense or utilitarianism are not undercut by hard incompatibilism per se (ibid, p115). This seems plausible, as these justifications for punishment are not outright grounded in the criminal being blameworthy. Similar to the striving for moral improvement I discussed in section 4.3, we can direct a similar attitude towards others in the face of wrongdoing. In the face of crime, this is a particularly important form of moral improvement, as a lack of moral improvement could result in further crimes. An argument that emphasizes the importance of moral improvement and prevention of further crime to justify punishment can thus be made without needing to assign blame to criminals.

In my view, Ferdinand Schoeman makes an even stronger case for justifying criminal punishment without assigning blame by drawing a comparison between the treatment of criminals and the carriers of diseases (Schoeman, 1979). Pereboom calls this the "Quarantine view", and summarizes Schoeman's comparison as follows:

> Ferdinand Schoeman (1979) contends that if we have the right to quarantine carriers of severe communicable diseases to protect people, then for the same reason we also have the right to isolate the criminally dangerous. Quarantining someone can be justified when she is not morally responsible for being dangerous to others. If a child is infected with a deadly contagious virus passed on to her prior to birth, quarantine may nevertheless be legitimate. Now suppose that a serial killer continues to pose a grave danger to a community. Even if he is not morally responsible for his crimes, it would be as legitimate to detain him as it is to quarantine a non responsible carrier of a deadly communicable disease. (Pereboom, 2007, p116)

I find this comparison appropriate, as it both says something about on what grounds we can justify detainment and how this detainment should be conducted. On the latter point, Pereboom writes that society benefits from a person with cholera getting quarantined, even though she does not merit this deprivation, but society should also do what it can to make the victim safe for release as soon as possible (Pereboom, 2001, p178). Building on this idea, Pereboom also points out that if we adopted the same attitude for detaining serial killers as carriers of diseases, it would also be morally wrong to treat those with criminal tendencies more severely than needed (ibid, p175).

It seems uncontroversial to me that treating carriers of diseases more severely than we need to do is morally wrong. I also hold that if we do not have moral responsibility, the quarantine view makes an intuitive comparison between criminals and carriers of diseases. Therefore, I hold that there would be good reasons to base justice systems on the quarantine view if hard incompatibilism holds true. Furthermore, adopting the quarantine view also retains certain aspects of our current justice system. Specifically, the quarantine view is robust against the argument that retributive justice is needed to disincentivize others from committing similar crimes. However, I believe that under the quarantine view, criminals would still be disincentivized from committing crimes. On this point, Pereboom writes that implementing the quarantine view would still have a preventive effect, as people would not desire to be detained, even if the detainment was conducted humanely and for the sake of preventive treatment (ibid, p177). This is a straightforward point, and I see no good reason why the risk of detainment would not have preventive effects on criminals. While detainment is inherently controversial due to depriving people of their freedom, I still argue that for the abovementioned reasons, the quarantine view provides a more justifiable approach to detainment than retributive justice does.

If the quarantine view was implemented, society would need to do what is possible to make criminals safe for release and show that rehabilitation can achieve this in a superior manner to a justice system focused on retribution. While there is an overlap between what constitutes rehabilitation and punishment (Ward, 2010), rehabilitation has, during the last 50 years, started to be seen in a more positive light and has successfully reduced recidivism.

> In 1979, he *(Robert Martinson)* acknowledged that under various conditions there are
> many examples of successful rehabilitative efforts . . . "such startling results are found

again and again . . . for treatment programs as diverse as individual psychotherapy, group counselling, intensive supervision, and what we have called individual help." About the same time, Paul Gendreau and Robert Ross reviewed ninety-five psychological intervention programs for criminals, and found that 86 percent were successful, with reductions in recidivism from 30 percent to 60 percent. The claims of their landmark study have been supported by more recent findings. (Pereboom, 2001, p181, my formatting and clarification in italics)

More recent data show similar results. In 1996, Henning and Frueh conducted an experiment on male offenders who partook in a cognitive-behavioral treatment program in state prison, finding a significant difference in recidivism between the group and a control group which did not partake in the program (Henning, Freuh, 1996). Similar to the treatment program in Henning & Frueh's work, the Risk-Need-Responsivity Model, which emphasizes cognitive social learning, has produced similar results.[34] Notably, in 2010, Donald J. Andrews and James Bonta wrote about how increasing punitive measures have failed to reduce recidivism and that programs that adhere to the Risk-Need-Responsivity Model have seen a reduction in recidivism by up to 35% (Andrews, Bonta, 2010). Therefore, at the very least, I find it reasonable to claim that quarantining for the sake of rehabilitation leads to moral improvement to a higher degree than quarantining purely for the sake of retributive punishment.[35]

I argue this reduction in recidivism is closely tied to the notion of moral improvement and showcases how moral responsibility is less significant for justice than one might intuitively believe. Another reason I make this claim is because attempting to refrain from some immoral action is merely a sufficient condition for avoiding being blamed and punished for it. In contrast, it is a necessary condition if one is to attain moral improvement. To illustrate why, consider how a child, around 13 years old, can opt to react to blame, and how they can opt to react to a

---

[34] Andrews and Bonta explain the Risk-Need-Responsivity Model the following way: The model describes: a) who should receive services (moderate and higher risk cases), b) the appropriate targets for rehabilitation services (criminogenic needs), and c) the powerful influence strategies for reducing criminal behaviour (cognitive social learning). (Andrews, Bonta, 2010, Abstract)

[35] Other forms of justification not grounded in retribution can be provided for imprisonment or punishment. Notably, there are theories based on the idea that the state can retaliate against offenders in virtue of removing advantages that are unfairly gained by refusal to obey the laws (Tebbit, 2017, p252). I believe that theories of this kind could be justified without assigning blame, and they could, therefore, be worthwhile to explore further as a hard incompatibilist.

discussion focused on moral improvement for some immoral action they did. Blame as a deterrence mechanism for preventing the child from doing some action again could prove successful, though not necessarily. If the child cannot understand why they did wrong, their motivation will be to avoid being blamed again, not to refrain from some action they do not themselves deem immoral. In turn, blame can be avoided by not getting caught when doing some such action again, lying about behaving in such a way, or through blame shifting or gaslighting. Therefore, refraining from doing some immoral action would be a sufficient condition for avoiding blame for some such action but not a necessary condition for avoiding blame for some such action. In turn, if the child wants to avoid blame, they can attempt to refrain from some such action or attempt to avoid blame in other ways. However, to attain moral improvement concerning whether they should do some immoral action or not, it is a necessary condition to attempt to refrain from doing some such action. If moral responsibility is assigned to the child, they may try to refrain from some such action, or they may not. If the child wants to attain moral improvement, the child must try to refrain from some such action.

In light of the above example and the reduction in recidivism in programs that emphasize rehabilitation, the problem of assigning moral responsibility can be viewed on a larger scale. In assigning moral responsibility, we raise children who may not attempt to refrain from immoral behavior. In turn, some of the children who continue to engage in immoral behavior will, as they grow older, engage in immoral behavior that is also punishable by law, leading to them getting incarcerated. In getting incarcerated, these people will again be blamed for their actions. They serve their sentence, an economic cost and an emotional toll for the criminal and their family. Then, they get released after having served their sentence. Again, refraining from criminal acts is a sufficient but not necessary condition for avoiding reincarceration. In not desiring to be reincarcerated, these people may or may not attempt to avoid criminal behavior in attaining this desire. However, if they strive for moral improvement, attempting to refrain from criminal acts is a necessary condition. Therefore, if moral improvement is the person's goal, they will attempt to avoid criminal behavior. If they do not strive for moral improvement, this increases the risk of them getting reincarcerated any number of times. At any point in time, from upbringing to incarceration to reincarceration, the focus could have been shifted from moral responsibility to moral improvement. As it was not, the focus on moral responsibility played a part in producing an outcome with severe costs on a personal and societal level.

In so far as detainment is concerned, the quarantine view could provide a solution to the abovementioned problem. I believe this has already successfully happened to some degree, as rehabilitation programs reflect some of the same values of the quarantine view. Therefore, adopting this view on a larger scale could have a profound benefit, and it would not require us to assign moral responsibility to criminals. The abovementioned subject also showcases why I deem moral responsibility to be a harmful notion. In not wanting to vanquish our sense of moral responsibility, we also assign blame and punishment. However, if we desire moral improvement, rehabilitation and safety, moral responsibility and blame is not a necessity.

I argue that the abovementioned points also highlight why we should not endorse an illusionist account on free will in favor of hard incompatibilism. Illusionist accounts claim that we ought to believe we have free will despite the lack of evidence, and the most prominent defender of this view is Saul Smilansky (Smilansky, 2000). Most notably for the present discussion, he writes that we are fortunate to be under the illusion that we have free will because we could not function well if we realized the full truth on the free will issue (Smilansky, 2000, p145). However, this is the account on free will that I have the most issues with. There seems little reason to do adopt an illusionist view unless a belief in moral responsibility and free will is essential or useful to us. I do not believe such a belief is essential for the reasons outlined in this chapter on hard incompatibilism. While moral responsibility might sometimes be useful, it can also be damaging, especially when moral responsibility is emphasized over moral improvement. I have also gone into detail why a lack of belief in us having free will would not have the type of severe consequences Smilnansky and others have claimed for the reasons outlined in sections 4.2 and 4.3. I have argued that we could benefit from disregarding or greatly revising our concept of moral responsibility and that doing so could have a positive personal, societal, and economic impact. Adopting an illusionist account on free will would prevent the possibility of this positive impact. Hard incompatibilism is the most plausible stance that also makes this positive impact possible.

Conclusion

In this thesis, I have argued that compatibilism does not hold true. I have outlined why by showing that Frankfurt cases fail to show how we can be said to be morally responsible if our actions are determined. This is because it would still be impossible for us to do anything but one predetermined action in any situation if determinism holds true, and Frankfurt cases fail to say anything meaningful about how the scenarios in Frankfurt cases differ from any everyday situation in this regard. Then, I argued that the consequence argument is a robust defense for why incompatibilism holds true. I have shown how one reason for this is that the argument can be upheld in its essentials, even if it requires reformulation. I have also shown how the truth of the premises and the validity of the argument are difficult to dispute. This was mainly done by showing how a hypothetical analysis of ability provides an unintuitive answer to how the consequence argument might be circumvented. The reason for this is that a hypothetical analysis of ability also bears with it the claim that we have the ability to perform actions that are impossible if determinism holds true. Therefore, any other argument that utilizes a hypothetical analysis of ability to attack the consequence argument will run into the same issue of having to claim that we can do something that cannot happen in this world.

After establishing why I support incompatibilism, I argued that libertarians have been unable to provide a convincing answer to how we can have free will and moral responsibility. If we had agent-causal powers, I would agree that we have free will, but I agree with Pereboom that it is unlikely that we do possess such powers. I agree with Kane and van Inwagen, who find the notion of agent causation mysterious. Older accounts of agent-causal libertarianism, such as Reid's and Chisholm's, make the claim that agents can cause events by agent-causing brain events. This essentially pushes the question of how we can cause our own brain events to occur to the scientists of the future. However, contemporary science seems to point to the contrary of what Reid and Chisholm have claimed.

While O'Connor's modern account of agent-causation is more refined than Chisholm and Reid's, this account runs into problems of its own. Notably, in attempting to account for the role of reasons, O'Connor's account invokes the question of how states can be said to cause anything. Most importantly, it is hard to see how O'Connor's claim that agent-causation should not be

abandoned but rather reconciled with modern science is a project that is worthy of pursuing. For all we know, agent-causation could be reconciled with science. However, as it seems unlikely that we have agent-causal powers according to modern science, this is, at the very least, a good reason to pursue other alternatives that seem more likely to hold true. Therefore, as it stands today, I argue that agent-causal libertarianism does not provide a plausible answer to how we can be said to have free will or be morally responsible.

I have also considered Robert Kane's event-causal libertarianism as an alternative. While his account answers questions compatibilism and agent-causal libertarianism struggle with, I have found that Kane's view also creates a wide range of problems of its own. As his view has many components that rely on each other to hold true, and many of these are controversial or implausible, I deem his view fragile and prone to Occam's razor. Most damning of these issues is his inadequate account of how our first Self-forming actions can come about and that his answer relies on endorsement, which I have shown to be a deeply problematic concept of its own. I praise Kane's attempt at grounding his view in natural sciences, which I believe is a step in the right direction for libertarianism. However, given that his view is based on unconfirmed and controversial theories, and his view hinges on multiple such theories holding true, I find Kane's account of event-causal libertarianism implausible. I give Kane credit for attempting to build libertarianism from the ground up, which I consider the right approach if one is to support this view. However, I find that he has been unsuccessful in providing a robust account of libertarianism and that it would be difficult to revise his account due to how intertwined the components of his view are.

In light of the major problems that face libertarianism and compatibilism, I find hard incompatibilism a more plausible and preferable alternative, as the position circumvents most of the issues libertarianism and compatibilism faces. True, many philosophers consider it undesirable to deny that we have free will. In answer to this, I critiqued van Inwagen's view that we cannot deliberate if we believe we only have one course open, making the claim that he is outright mistaken in this regard. I also expanded on the theories of Derk Pereboom to show why hard incompatibilism would not be an undesirable position in other important regards. I argued that we can still retain many of the notions that make up our interpersonal relationships, such as gratitude and love. Then, I argued that even if certain forms of blame and praise would not be

justified, it would not hinder moral improvement and that we could not only retain a justice system but improve upon it to be more pragmatic and just on a hard incomaptibilist view.

A vital point of this argumentation has been to show that the free will discourse would benefit from more thoroughly considering the plausibility of hard incompatibilism. I find that hard incompatibilism provides a far more plausible answer to the question of whether we have free will than libertarianism and compatibilism do. One area where this is especially apparent is the impact of the consequence argument if determinism turned out to be true or false. I hold that the strength of the consequence argument favors hard incompatibilism even more than libertarianism, as libertarianism hinges on determinism not holding true, whereas hard incompatibilism does not. If the consequence argument is valid, and determinism turned out to hold true, this would render libertarianism false, but hard incompatibilism could still hold true. If determinism does not hold true, premise 2 of the consequence argument, "No one has power over the fact that the facts of the past and the laws of nature entail every fact of the future" would be false. In this case, either libertarianism or hard incompatibilism could hold true. Therefore, this robust argument will always be compatible with hard incompatibilism, but it could turn out not to be viable as a defense of libertarianism, as it could instead disprove the position outright.

In chapters 1, 2, and 3, I highlighted how positions that hold that we have free will have gone to great lengths to attempt to explain how we can be said to have free will and be morally responsible. Compatibilists have used a hypothetical analysis of ability to claim that we can perform impossible actions. Multiple mystic accounts of how agent-caused actions can occur have been provided in defense of agent-causal libertarianism. First self-forming actions have been claimed to be sufficiently free for character formation due to the endorsement of one outcome on Kane's event-causal libertarian view. To me, the implausibility of these claims and others that have been discussed far outweighs the implausibility of us simply not having free will. Hard incompatibilism raises questions of its own, but in terms of plausibility, I hold that we have significantly better reasons for holding this view than supporting libertarianism or compatibilism.

For these reasons, I claim that the presumption that a lack of free will is undesirable has a negative impact on the free will debate. This presumption has been expressed as both an undesirability for adopting skeptical positions, and the view that skeptical positions holding true

would be undesirable. However, I argue that hard incompatibilism is not undesirable to believe in and it would not be an undesirable outcome if this position held true. Neither of these presumptions are, as I have argued in sections 4.2, 4.3 and 4.4, as well-founded as many philosophers have claimed. I, therefore, also argue that many thinkers could benefit from reexamining and revising their views on this matter and that not doing so can also have negative effects on their views on justice, blame and praiseworthiness.

By showing why hard incompatibilism is not an undesirable position, my goal has been to illuminate what I consider to be a common misunderstanding about skeptical positions. As I have shown in section 4.4, whether we have free will and moral responsibility significantly impacts our lives. I believe the best approach possible is to focus our attention on finding the most plausible answer to whether we have free will or not. If the undesirability of some position holds philosophers back from giving it serious consideration, this can be a hindrance to uncovering important truths about the subject. This is not merely detrimental to how well philosophers can approach the metaphysics of free will, but this also has an impact on our everyday lives to various degrees. Given how impactful our notions of free will and moral responsibility are to our lives, no view should be neglected because it is deemed undesirable. The consequences of making moral judgments on an incorrect or unconvincing account of free will are what I deem most undesirable of all.

In conclusion, I hold that compatibilism and libertarianism provide lacking answers to the question of whether we have free will and moral responsibility. Hard incompatibilism provides a more plausible alternative to this question by avoiding the problems that these positions face. Hard incompatibilism is also more desirable than many philosophers have argued, both in terms of adopting the view and the view holding true, which could motivate some to consider the position more seriously. By abolishing our sense of moral responsibility, we can make more just decisions on an interpersonal and societal level. Reflecting on the desirability of the position could be productive to the philosophical debate surrounding free will, ensuring that arguments are more well-founded by not presupposing an unfounded undesirability for the position. Most importantly, hard incompatibilism is worthy of consideration due to the impact the truth value of free will and moral responsibility have on our personal and societal lives. I have shown why I hold this to be the case and why disregarding hard incompatibilism can have everything from

minor to severe consequences on how we treat others and how society treats criminals. In this regard, hard incompatibilism offers both the most plausible and least damaging account, and I claim that neglecting to consider it as an alternative can negatively impact people's lives on a multitude of levels.

Bibliography

Alvarez, M. (2009). Actions, thought-experiments and the 'principle of alternate possibilities'. Australasian Journal of Philosophy 87 (1), p61–81. DOI:10.1080/00048400802215505

Andrews, D. A., Bonta, J. (2010). Rehabilitating Criminal Justice Policy and Practice, in *Psychology, Public Policy, and Law* 2010, Vol. 16, No. 1, p39–55. DOI:10.1037/a0018362

Austin, J. L. (1956). Ifs and Cans, *Proceedings of the British Academy* 42: 109–132; reprinted in 1979, p205–232. doi:10.1093/019283021X.003.0009

Basar, E.(ed) (1990). *Chaos in Brain Functions*, Berlin, Springer-Verlag, DOI:10.1007/978-3-642-75545-4

Blumenfeld, D. (1971). The Principle of Alternate Possibilities, *The Journal of Philosophy* vol 68 (11), p339-345, DOI:10.2307/2024867

Carr, D. D., (2019). Counterfeit Indeterminacy and Kane's Self-Forming Actions. *Southwest Philosophy Review* 35 (1), p99-106. DOI:10.5840/swphilreview201935110

Caruso, Gregg D. (2015). Kane is Not Able: A Reply to Vicens' "Self-Forming Actions and Conflicts of Intention". *Southwest Philosophy Review* 31 (2), p21-26. DOI:10.5840/swphilreview201531231

Caruso, G. D. (2018). Skepticism about Moral Responsibility, *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/skepticism-moral-responsibility/

Chalmers, D., Bourget, D. (2023). Philosophers on Philosophy: The 2020 PhilPapers Survey. *Philosophers' Imprint,* 0(0). DOI:10.3998/phimp.2109

Chisholm, R. M. (1964) Human Freedom and the Self. *The Lindley Lecture*, Department of Philosophy, University of Kansas, April 23, 1964, p3-15, http://hdl.handle.net/1808/12380

Clarke, R. (1993). Toward A Credible Agent-Causal Account of Free Will, *Noûs*, 27(2), p191–203. https://doi.org/10.2307/2215755

Clarke, R. (2003). *Libertarian Accounts of Free Will*. New York, US: Oxford University Press USA.

Clarke, R., Capes, J., Swenson, P. (2021). Incompatibilist (Nondeterministic) Theories of Free Will, *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/incompatibilism-theories/

Copp, D. (1997). Defending the principle of alternate possibilities: Blameworthiness and moral responsibility, *Noûs* 31 (4), p441-456. DOI:1111/0029-4624.00055

Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. London, England: MIT Press.

Díaz, M. H., Córdova, F. M., Cañete, L., Palominos, F., Cifuentes, F., Sánchez, C., Herrera, M. (2015) Order and Chaos in the Brain: Fractal Time Series Analysis of the EEG Activity During a Cognitive Problem Solving Task, *Procedia Computer Science*, Volume 55, p1410-1419, https://doi.org/10.1016/j.procs.2015.07.135.

Dorr. C, (2016) Against Counterfactual Miracles, *Philosophical Review* vol. 125, no. 2 (2016), 241–286. DOI:10.1215/00318108-3453187

Epicurus. (2014). Letter to Menoeceus, *The Essential Epicurus.* The Big Nest.

Falkenburg, B., Weinert, F. (2009). Indeterminism and Determinism in Quantum Mechanics, Greenberger, D., Hentschel, K., Weinert, F. (eds) *Compendium of Quantum Physics*. Springer, Berlin, Heidelberg, p307-311, DOI:10.1007/978-3-540-70626-7_96

Feldman, R. H. & Buckareff, A. A. (2003). Reasons explanations and pure agency, *Philosophical Studies* 112 (2), p135-145. DOI:10.1023/a:1022542004219

Fischer, J. M., Kane, R., Pereboom, D., Vargas, M. (2007) *Four Views on Free Will*. Blackwell Publishing

Fischer, J. R. (2010). The Frankfurt cases: The moral of the stories, *Philosophical Review* 119 (3), p315-336. DOI:10.1215/00318108-2010-002

Fischer, J. M. (2012). Semicompatibilism and Its Rivals, *The Journal of Ethics,* 16(2), 117-143. DOI:10.1007/s10892-012-9123-9

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility, *Journal of Philosophy* 66 (23), p829-839. DOI:10.2307/2023833

Franklin, C. E. (2014) Event-causal libertarianism, functional reduction, and the disappearing agent argument, *Philos Stud* 170, p413–432. DOI:10.1007/s11098-013-0237-0

Ginet, C. (1966). Might We Have No Choice, K. Lehrer (Ed.), *Freedom and Determinism.* Random House, p87-104.

Glatz, R. M. (2008). The (near) necessity of alternate possibilities for moral responsibility, *Philosophical Studies* 139 (2), p257-272. DOI:10.1007/s11098-007-9116-x

Graham, P. A. (2008). A Defense of Local Miracle Compatibilism, *Philosophical Studies*, 140, p65–82. DOI:10.1007/s11098-008-9226-0

Griffith, M. (2010). Why Agent-Caused Actions Are Not Lucky, *American Philosophical Quarterly*, 47(1), 43-56. https://www.jstor.org/journal/amerphilquar

Harrison, G. (2011). Frankfurt's Refutation of the Principle of Alternative Possibilities, Bruce, M. & Barbone, S. (eds.), *Just the Arguments*. Wiley-Blackwell. p121–122. DOI:10.1002/9781444344431.ch31

Henden, E. (2010). Deliberation Incompatibilism, *Dialectica* 64 (3):313-333. DOI: 10.1111/j.1746-8361.2010.01234.x

Henning, K. R., Frueh, B. C. (1996). Cognitive-Behavioral Treatment of Incarcerated Offenders: An Evaluation of the Vermont Department of Corrections' Cognitive Self-Change Program, *Criminal Justice and Behavior*, 23(4), p523-541. DOI:10.1177/0093854896023004001

Hoefer, C. (2023). Casual Determinism, *Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/determinism-causal/

Jaeger, G. (2017). Quantum randomness and unpredictability, *Fortschr. Phys*. 2017, 65, 1600053. https://doi.org/10.1002/prop.201600053

Jaworski, W. (2011). *Philosophy of Mind: A Comprehensive Guide*, Wiley-Blackwell

Kane. R, (1996). *The Significance of Free Will*. Oxford University Press, New York

Kane. R, (2002). Some Neglected Paths in The Free Will Labyrinth, Kane, R. (ed.) *The Oxford Handbook of Free Will*, Oup Usa, (p406-440)

Kane. R, (2005). *A Contemporary Introduction to Free Will. New York*. Oxford University Press.

Kapitan, T. (2002) A Master Argument For Incompatibilism, Kane, R. (ed.) *The Oxford Handbook of Free Will*, Oup Usa, (p127-158)

Kraut, R. (2022), Aristotle's Ethics, *Stanford Encyclopedia of Philosophy,* https://plato.stanford.edu/entries/aristotle-ethics/

Lamb, J. W. (1993). Evaluative Compatibilism and the Principle of Alternate Possibilities, *Journal of Philosophy* 90 (10), p517-527.

Lemos, J. (2015). Self-Forming Acts and the Grounds of Responsibility, *Philosophia* 43, p135–146. https://doi.org/10.1007/s11406-014-9561-7

Lewis, D. (1981). Are We Free to Break the Laws?, *Theoria*, 47 (3), p113–121. DOI:10.1111/j.1755-2567.1981.tb00473.x

Mandik, P. (2013). *This is Philosophy of Mind: An Introduction*. Hoboken, Wiley-Blackwell.

Markosian, N. (1999). A Compatibilist Version of the Theory of Agent Causation, *Pacific Philosophical Quarterly*, 80(3), 257-277. DOI:10.1111/1468-0114.00083

McKay, T. J. & Johnson, D. (1996). A Reconsideration of an Argument against Compatibilism, *Philosophical Topics* 24 (2), p113-122. DOI:10.5840/philtopics199624219

McKenna, M., Coates, D. J. (2024). Compatibilism, *Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/compatibilism/

Mele, A. R. (2006). *Free Will and Luck*. New York, US: Oxford University Press.

Murday, B. (2017). Can Self-Forming Actions Dispel Worries about Luck?, *Philosophia* 45 (3), p1313-1330. DOI:10.1007/s11406-016-9808-6

Myvold, W. (2022). Philosophical Issues in Quantum Theory, *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/qt-issues/

Naylor, M. (1984). Frankfurt on the principle of alternative possibilities, *Philosophical Studies* 46 (September), p249-58. DOI:10.1007/bf00373108

Nichols R. Yaffe G. (2014), Thomas Reid, *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/reid/

Oakley, S. (2006). Defending Lewis's Local Miracle Compatibilism, *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 130 (2), p337–349. DOI:10.1007/s11098-004-4677-4

O'Connor. T., (1996). Why Agent Causation?, *Philosophical Topics* 24 (2), p143-158. DOI:10.5840/philtopics199624218

O'Connor, T. (2000). *Persons and Causes: The Metaphysics of Free Will*. New York, US: Oxford University Press USA.

O'Connor, T., Franklin, C. (2022). Free will, *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/index.html

Paterek, T., Kofler, J., Prevedel, R., Klimek, P., Aspelmeyer, M., Zeilinger, M., Brukner C. (2010). Logical independence and quantum randomness — with experimental data, *New J. Phys.* 12, 013019, DOI:10.1088/1367-2630/12/1/013019

Pendergraft, G. (2011). The Explanatory Power of Local Miracle Compatibilism, *Philosophical Studies* 156 (2), p249-266. DOI:10.1007/s11098-010-9594-0

Pereboom, D. (2001). *Living Without Free Will.* Cambridge University Press.

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford University Press.

Pereboom, D. (2017). Responsibility, Agency, and the Disappearing Agent Objection, Guillon J.-B. (Ed.), *Le libre arbitre* (1–). Collège de France. DOI:10.4000/books.cdf.4942

Perry, J. (2003). *Is There Hope for Compatibilism?* David Norton Lecture, University of Delaware. April 21, 2003. http://john.jperry.net/cv/2004cHopeForCompat.pdf

Perry, J. (2010). Wretched subterfuge: a defense of the compatibilism of freedom and natural causation, *Proceedings and Addresses of the American Philosophical Association* 84 (2), p93-113. http://www.jstor.org/stable/25769937

Potter, H. D., Mitchell K. J. (2022) Naturalising Agent Causation, *Entropy*; 24(4), p472. https://doi.org/10.3390/e24040472

Robb, D. (2020), Moral Responsibility and the Principle of Alternative Possibilities, *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/alternative-possibilities/

Schoeman, F. D. (1979). On Incapacitating the Dangerous, *American Philosophical Quarterly*, 16(1), p27–35. http://www.jstor.org/stable/20009736

Skarda, C. A., Freeman, W. J. (1987). How brains make chaos in order to make sense of the world, *Behavioral and Brain Sciences*, 10 (2), p161–195. DOI: 10.1017/S0140525X00047336

Skarda, C. A., Freeman, W. J. (1990). Chaos and the new science of the brain, *Concepts in Neuroscience,* Vol. 1, No. 2, p275–285, World Scientific Publishing Company

Smilansky, S. (2000). *Free Will and Illusion*. New York, Oxford University Press

Strawson, G. (1994). The impossibility of moral responsibility, *Philosophical Studies* 75 (1-2), p5-24. DOI:10.1007/bf00989879

Strawson, P. (1962). Freedom and Resentment, Fischer, J. M., & Ravizza, M. (eds.), *Perspectives on Moral Responsibility*. Cornell University Press, p1-25. DOI:10.7591/9781501721564-002

Talbert, M. (2019). Moral Responsibility, *Stanford Encyclopedia of Philosophy,* https://plato.stanford.edu/entries/moral-responsibility/

Tebbit, M. (2017) *Philosophy of Law An Introduction*, New York, Routledge

Van Inwagen, P. (1975). The Incompatibility of Free Will and Determinism, *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 27(3), p185–199. DOI:10.1007/bf01624156

Van Inwagen, Peter (1983). *An Essay on Free Will*. New York, Oxford University Press.

Van Inwagen, P. (2002) Free Will Remains a Mystery, Kane, R. (ed), *The Oxford Handbook of Free Will*, Oxford University Press, New York

Van Inwagen, P. (2008). How to Think about the Problem of Free Will, *The Journal of Ethics* 12 (3-4), p327-341. DOI:10.1007/s10892-008-9038-7

Van Inwagen, P. (2015), *Metaphysics* (4th ed), Boulder Colorado, Routledge

Vicens, Leigh C. (2015). Self-Forming Acts and Conflicts of Intention, *Southwest Philosophy Review* 31 (1), p93-100. DOI:10.5840/swphilreview20153119

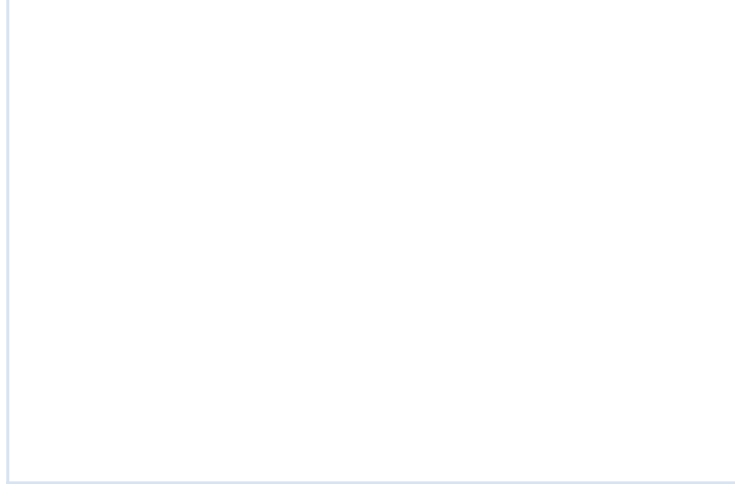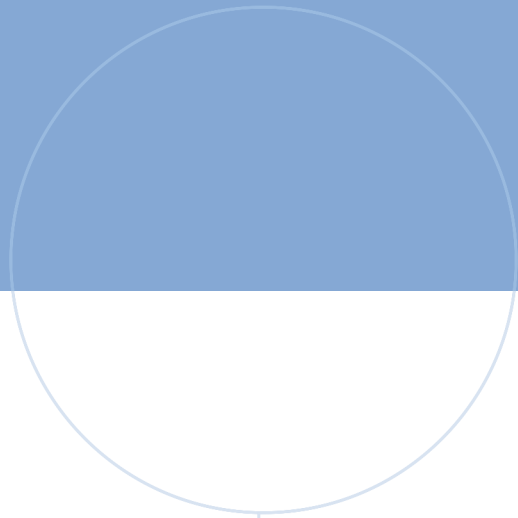Vihvelin, K. (2022). Arguments for Incompatibilism, *Stanford Encyclopedia of Philosophy,* https://plato.stanford.edu/entries/incompatibilism-arguments/

Wakefield, T. (2003) Compatibilism and Incompatibilism: Some Arguments, Loux, M. J., & Zimmerman, D. W. (eds.). *The Oxford handbook of metaphysics.* New York: Oxford University Press.

Ward, T. (2010), Is offender rehabilitation a form of punishment?, *The British Journal of Forensic Practice*, vol 12 No4, p4-13 DOI:10.5042/bjfp.2010.0610

Widerker, D. (2005), Agent-Causation and Control, *Faith and Philosophy* Volume 22, Issue 1, p87-98, DOI:10.5840/faithphil200522139

Williams, B. A. O., Atkinson W. F. (1965). Symposium: Ethical Consistency, *Proceedings of the Aristotelian Society*, Supplementary Volumes, 39, p103–138. http://www.jstor.org/stable/4106568

Wolf, S. (1987). Sanity and the Metaphysics of Responsibility, Schoeman, F. D. (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology.* Cambridge University Press, p46-62.