2013-04-29

No Fate But What We Make - A Defense of the Compatibility of Freedom and Causal Determinism

Ryan N. Lake
University of Miami, r.lake@umiami.edu

UNIVERSITY OF MIAMI

NO FATE BUT WHAT WE MAKE – A DEFENSE OF THE COMPATIBILITY OF FREEDOM AND CAUSAL DETERMINISM

By

Ryan N. Lake

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2013

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

NO FATE BUT WHAT WE MAKE – A DEFENSE OF THE COMPATIBILITY OF FREEDOM AND CAUSAL DETERMINISM

Ryan N. Lake

Approved:

_____ Michael Slote, Ph.D. UST Professor of Ethics	_____ M. Brian Blake, Ph.D. Dean of the Graduate School
_____ Bradford Cokelet, Ph.D. Assistant Professor of Philosophy	_____ Risto Hilpinen, Ph.D. Professor of Philosophy

_____
Keith Lehrer, Ph.D.
Regents Professor of Philosophy, Emeritus
University of Arizona

LAKE, RYAN N. (Ph.D., Philosophy)
No Fate But What We Make – A Defense of the Compatibility of Freedom and Causal Determinism (May 2013)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Michael Slote.
No. of pages in text. (190)

In this dissertation I explore the question of the compatibility of freedom and moral responsibility with causal determinism. A number of philosophers and thinkers have argued that if causal determinism were true, that our ordinary attributions of free will and responsibility would be completely undermined. I argue that this claim is ultimately mistaken, and that there are robust and common sense notions of freedom and responsibility that are applicable even if everything we do is ultimately causally determined. I start by building a general framework for understanding freedom and moral responsibility from the standpoint of practical reason that incorporates moral reactive attitudes, and in part by using this framework, I develop detailed replies to the most compelling and powerful arguments in favor of incompatibilism that have been developed in recent decades, most notably in the work of philosophers like Derk Pereboom and Bruce Waller.

For Bunny, with eternal love

Acknowledgements

I'd like to thank my committee members for their tremendous help and encouragement – Michael Slote, Risto Hilpinen, Brad Cokelet, and especially Keith Lehrer. It was Keith's seminar on Free Will in my first semester of grad school that piqued my interest in this topic and made me realize I might have something to say about it, and I have enjoyed and benefited from our interactions over the years immensely. I want to thank many supportive friends, including Ben Burgis, for helpful conversations and for reading early drafts of my work, Sarah Lesson, for the countless hours she spent trying to convince me that my views about free will are wrong, Mark Warren, for coercing me into wagers that forced me to write large portions of this dissertation, and all of the Huffles (Micah Dugas, Mark Warren, Stephanie Saline, Ben Yelle, Robin Neiman, and Fredrik Haraldsen) for providing me with a real home in Miami over the last few years. I want to thank my wonderful and supportive family, especially my mom Cathy Schultz, who instilled in me a passion for reading and learning from an early age, and who always encouraged me, even when I decided to do something as crazy as majoring in philosophy. I also want to thank Blair Morrissey, my first philosophy professor, for sparking my passion for this subject, for having so much confidence in me from the beginning, and for years of rewarding and enjoyable conversations. And I would especially like to thank my sweetheart, Bunny Sandefur. Her unfailing love and support (not to mention her incredible patience with me) as I went through the challenging process of finally pulling this project together made it all possible, and I will be forever grateful.

Table of Contents

Chapter 1: Introduction
1

Chapter 2: Moral Responsibility and Practical Reason
8

Chapter 3: Leeway Incompatibilism
45

Chapter 4: Source Incompatibilism
84

Chapter 5: Foreknowledge and Moral Responsibility
124

Chapter 6: The Importance of Moral Responsibility
151

Bibliography
184

"There is no fate but what we make for ourselves."
–John Connor, "Terminator 2: Judgment Day"

Chapter 1 - Introduction

This quotation contains one of the most memorable lines from the Terminator series, a sort of mantra expressing a central philosophy of the Terminator franchise. Interestingly, the line "no fate but what we make" was supposed to be uttered by Kyle Reese in the first Terminator film. But it was cut out, shortened merely to "the future is not set". It might seem that this is meant as a metaphysical claim, a claim that there are simply no facts about what will happen in the future. And certainly the later additions to the Terminator franchise (the TV series "The Sarah Connor Chronicles, the later films, etc.) seem to interpret things in this way, as John and Sarah Connor (along with help from people and cyborgs from the future) do things that substantially alter the way the future history plays out. In the second and third films we learn that their actions have substantially delayed "Judgment Day" - the day that the self-aware computer network Skynet rains nuclear destruction on humanity - and their aim throughout the TV series was to make sure it never happened at all.

But what if this wasn't true - what if there are facts about the future? What if Judgment Day was fixed, unalterable, and it was a fact (at the times of their actions in the earlier films) that everything that Kyle, Sarah, and John actually do will ultimately lead to Judgment Day occurring on August 29th 1997 (the first date given in the Terminator films)? Would this mean that Kyle (and ultimately John Connor of the future) was wrong to claim that we make our own fates? I will argue that on a clear, obvious, and intuitive understanding of this claim that it does not. This is why it was interesting that the quote

"no fate but what we make" was originally written for the script of the first movie, because the first film - unlike the others - seems to very explicitly have this unchangeable model of time in mind. For instance, we learn that Kyle Reese, the soldier sent back in time to save Sarah, is actually John's father. From this it follows that it must have already been true that John would send Kyle back in time before he actually did, because John's existence depends on that fact. We also see in flashbacks that Kyle has long possessed a photograph of Sarah - a gift given to him by John in the future. When Sarah gets her photograph taken by a child at a gas station at the end of the film, it is the exact same photograph that Kyle possessed. This causal loops also implies that it was already a fact that Kyle would be sent into the past and that the events of the film would play out exactly as they did before Kyle ever went back. And finally (though we don't learn it until the second film), the Terminator Machine from the future that is destroyed at the end of the first film ends up being the basis for the development of the technology that leads to Skynet, and ultimately to the development of the Terminator itself. This implies that it must have already been a fact that the Terminator would be sent back prior to its actually being sent back, because otherwise it would never exist at all.

The same general point can be made about all of the causal loops that occur in the Terminator series (which is why the alterable timeline metaphysics of the later entries in the franchise become difficult to make sense of). So we might conclude that James Cameron simply wasn't thinking carefully when he penned that pivotal phrase for the first script. But perhaps Cameron had a deeper insight in mind - perhaps the claim that "there is no fate but what we make" can be true even if there are set facts about what will happen in the future. That, in essence, is the claim I will be defending in this dissertation.

I will be arguing that the existence of predetermined facts about every event that occurs is perfectly compatible with the future being ours to make, in a deep, genuine, and intuitive sense. We can act freely and responsibly, shaping our lives in a ways that are up to us, even if there happen to be preset, causally determined facts about the ways in which we will actually go about doing that.

The main focus of this dissertation, then, will be defending the compatibility of causal determinism and freedom. Causal determinism is, in short, the claim that every event that occurs in the universe - including our own decisions and actions - is causally necessitated by earlier events (due to the laws of physics and the actual state of the universe at earlier times). Causal determinism is one way of having there be facts about what happens in the future, but it is not the only way. It's also entirely possible that the world is causally indeterministic and yet that a 4-dimensional or "block" theory of time is the correct one, meaning that the future is every bit as real as the present. I will explore that sort of possibility in a later portion of the dissertation, arguing that the mere existence of future facts is less threatening to freedom and responsibility than causal determinism itself is.

To begin with, I would like to be clear at the outset just what concept of freedom I am talking about. There are a number of different notions of freedom that different thinkers have been interested in. Perhaps the most basic notion of freedom is the sense that implies a lack of immediate physical restraint or immediate physical coercion. It is the freedom to act on one's desires and preferences, the freedom to live the sort of life that one wishes to live. I don't wish to downplay the significance of this sort of freedom. This is a sense of freedom that many people have fought and died for, and it is a notion of

freedom that frames many vexing debates in politics today (see, for example, current debates on gun control, taxation, reproductive rights, etc). Despite its great importance, this sense of freedom is not interesting when it comes to the question of its compatibility with causal determinism. There is little question about whether this variety of freedom is compatible with causal determinism; it obviously is. It's clearly the case that a person could be free in the sense of not being impeded from acting on his desires even if it were the case that all of his desires were causally necessitated by earlier events.

What I am interested in - and what most writers on the topic have been primarily interested in - is the sense of freedom required for moral responsibility. I will discuss the notion of moral responsibility in more detail shortly, but roughly to say that an agent is morally responsible is to say that he or she can rightly be praised or blamed for his or her actions. It has seemed to many people that something **more** than an absence of impediments to acting on one's desires is needed to ground moral responsibility. For instance, a common and intuitive assumption has been that in order to be truly free in the sense required for moral responsibility, an agent must have the ability to select among or choose among alternative courses of action – to do otherwise than one actually does. Some have referred to this as the "Garden of Forking Paths" model of freedom[1]. If this is the kind of freedom required for moral responsibility, then it is obvious why causal determinism poses a threat. For it seems that causal determinism tells us that (in some strict sense) only one future is ever possible - which implies that for any choice that an agent makes, only one alternative is ever really open to that agent. Thus it seems that

[1] See for example Fischer, John Martin. *The Metaphysics of Free Will: An Essay on Control*. Cambridge, MA: Blackwell, 1994.

prima face we have a good argument for thinking that freedom (in the deep sense we require for moral responsibility) is not compatible with causal determinism. Arguing against this prima facie appearance will be the main task of this dissertation. Here then is a rough outline of structure of this dissertation.

The **second chapter** will be dedicated to setting up the problem, first by exploring the concept of moral responsibility in considerable detail. From there, I will move into developing an account of moral responsibility that starts by looking at the contrasts between competing standpoints. In this I will draw on the work of several philosophers, most notably Hilary Bok, who argues that an understanding of free will and moral responsibility can be developed from the standpoint of practical reason. From that work, I will develop a framework that will be used to support a generally compatibilist viewpoint, and to explore and respond to a variety of arguments against compatibilism about moral responsibility and causal determinism in the later chapters.

The **third chapter** will be primarily focused on one way of arguing for the incompatibilist viewpoint, and will defend a major strategy of response to that argument. This way of arguing for incompatibilism appeals to the claim mentioned above – that moral responsibility requires the ability to do otherwise than one actually does – and has become known as "leeway incompatibilism".[2] The strategy I will defend is a line of response to leeway incompatibilism first developed by Harry Frankfurt[3] in which he argues that we can be morally responsible for our actions even if we lack the ability to anything other than what we actually do. I will critically survey some of the large volume

---

[2] Derk Pereboom, *Living without Free Will* (Cambridge, U.K.: Cambridge University Press, 2001), pg. 5-6.
[3] Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, no. 23 (1969): 829-839.

of debate that has grown around Frankfurt's argument. And I will focus especially on a recent significant and incisive line of attack on the Frankfurt strategy that has been advanced by Kadri Vihvelin[4] (in which she argues that Frankfurt examples commit a modal fallacy), and develop a line of response to her critique in part by drawing on the framework developed in the first chapter.

The **fourth chapter** will turn to a different way of motivating incompatibilism, which is has been called "causal history incompatibilism",[5] or sometimes alternately "source incompatibilism". According to this version of incompatibilism, the primary explanation of why moral responsibility is incompatible with causal determinism is that determinism means that we are not truly the sources of our own preferences and desires and decisions. A method of advancing this version of incompatibilism involves the appeal to manipulation cases – thought experiments in which agents satisfy all of the conditions of traditional compatibilist accounts, but are intuitively not responsible because they have been manipulated by an outside agent. I will particularly focus on a well-known version of this argument developed by Derk Pereboom.[6]

In the **fifth chapter**, I turn to some arguments related to the role that *time* plays in the debate over free will and moral responsibility. For instance, I respond to a recent argument against compatibilism advanced by Saul Smilansky that is based on the notion of "prepunishment",[7] or punishing people for crimes that they have not yet committed.

---

[4] Kadri Vihvelin, "Freedom, Foreknowledge, and the Principle of Alternate Possibilities", *Canadian Journal of Philosophy* 30, no. 1 (March 2000): 1-23.
[5] Pereboom, *Living without Free Will*, pg. 5-6.
[6] See Pereboom's discussion of the "four case" argument in *Living Without Free Will*, especially in Chapter 4.
[7] Saul Smilansky, "Determinism and Prepunishment: The Radical Nature of Compatibilism." *Analysis* 67, no. 4 (2007): 347-49.

Smilansky argues that compatibilism allows no room for a principled objection to prepunishment, and concludes that this shows that determinism has radically revisionary results for our ordinary moral concepts – contrary to what compatibilism claims. In response, I will argue that prepunishment is not just a problem for compatibilists (prepunishment problems can be generated for libertarians as well), and develop a way of resisting prepunishment that draws on the framework developed earlier.

The **final chapter** will attempt to bring the ideas of the earlier chapter together into a single, coherent picture. The framework advanced in the second chapter, wherein freedom and responsibility are understood in terms of the standards of practical reason, will be developed further in light of the arguments of the previous chapters. I will develop an account of the importance of moral responsibility to our moral and social lives, and I will also address some lingering skeptical concerns concerning the potential harms of our belief in moral responsibility. Ultimately I will argue that there should be a strong prima facie presumption in favor of the compatibilist viewpoint. Given the inadequacy of incompatibilist arguments, as demonstrated in the earlier chapters, I will conclude that we are well justified in accepting the truth of compatibilism about freedom and moral responsibility

Chapter 2 – Moral Responsibility and Practical Reason

The central concern of this dissertation is the question of whether the sort of freedom required for moral responsibility can exist in a causally deterministic world – in other words, whether moral responsibility is compatible with causal determinism. To get clear about what this question is asking, I want to begin first by considering the concept of moral responsibility. Just what does it mean to say that someone is morally responsible for an action or decision? A number of philosophers and thinkers have attempted to explain, or at least shed light on, the concept of moral responsibility by alluding to other related and more familiar concepts of responsibility. While these concepts are distinct, they are also in many ways closely related. I will consider a few of these concepts of responsibility in turn.

I. Attributability

One notion of responsibility that is frequently discussed is responsibility in the sense of attributability. The idea here would be that you are responsible for an action or a state of affairs if it can be causally attributed to you as an agent. This is closely analogous to a common sense of responsibility we use in every day life – for example, we might also say that a tornado is responsible for the destruction of a neighborhood in Missouri, or that a drought is responsible for a famine.

While this is a familiar sense of responsibility, and while it seems necessary for moral responsibility (it would seem inappropriate to say that someone is morally responsible for something that cannot be attributed to him), I don't think it is sufficient to fully explain moral responsibility on its own. On the face of it, it seems that sometimes people can be responsible for actions in the sense of attributability without clearly being

morally responsible those actions. For example, imagine a woman who is a kleptomaniac, and impulsively swipes a candy bar from a shelf. The theft can be attributed to her – the action can be traced to an aspect of her character, it reveals something about her[8] - but we still may conclude that she was not morally responsible. A full theory of moral responsibility needs to explain why we hold people accountable for *some* actions that can be attributed to them and not others. I will have a bit more to say about this concept of responsibility later in the chapter when I contrast it with the account that I prefer.

II. Punishment

Another notion that is often connected to the notion of moral responsibility concerns punishment and rewards. It might be thought that to be morally responsible for an action just means that you ought to be or deserve to be punished or rewarded for it (depending on whether it was a morally good or bad action), or at least that a punishment or reward would be appropriate.[9] Many people discuss moral responsibility in just this way.[10] According to Michael McKenna, "what most everyone is hunting for … is the sort of moral responsibility that is desert entailing, the kind that makes blaming and punishing as well as praising and rewarding justified."[11] McKenna then goes on to warn against framing the debate in this way. Similarly, Bruce Waller writes: "The moral responsibility that is my target is the moral responsibility that justifies special reward and punishment.

---

[8] For an example of this characterization of responsibility as attributability, see Gary Watson, "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227-48.
[9] See J.J.C. Smart "Free-Will, Praise And Blame." *Mind* LXX, no. 279 (1961): 291-306.
[10] For another recent example, see Sam Harris, *Free Will*, New York: Free Press, 2012.
[11] Michael McKenna. "Compatibilism & Desert: Critical Comments on "Four Views on Free Will"" *Philosophical Studies* 144, no. 1 (2009), 12.

Moral responsibility provides the moral justification for singling an individual out for condemnation or commendation, praise or blame, reward or punishment."[12]

While I agree that the concept of moral responsibility should certainly *inform* our practices of punishments and rewards, it seems clear to me that the concept of moral responsibility and the practice of punishing and rewarding are distinct. First, it is clear that being morally responsible for an action is not sufficient for saying that a punishment or reward is appropriate. There are many kinds of actions that we take to be immoral without thinking that they warrant any punishment. For instance, I may promise a friend that I will meet him for a drink after work, but then ditch him (without notifying him) when a girl that I am interested in invites me out for a drink. This is clearly an immoral action, but saying that I deserve any sort of punishment (especially legal punishment) for this action seems out of line.

Further, I think it can be plausibly argued that being morally responsible is also not *necessary* for saying that a punishment or reward is appropriate. Many philosophers have disagreed. For example Waller says, ""moral responsibility" is the essential (necessary, if not sufficient) condition for justified blame and punishment."[13] However this strikes me as a mistake. There seem to be at least some cases where it is appropriate to reward or punish someone even as we acknowledge that they are not really morally responsible for whatever it is that they are being punished or rewarded for. Most would acknowledge that very young children, for example, are not really morally responsible for their actions; they lack a sufficiently developed concept of right and wrong, they lack a sense of the consequences of their actions, and they lack a well developed capacity to

---

[12] Bruce Waller, *Against Moral Responsibility* (Cambridge, MA: MIT Press, 2011), 1-2.
[13] Waller, *Against Moral Responsibility*, 2.

restrain their impulses, all of which seem to be important conditions for moral responsibility (whether one is a compatibilist or an incompatibilist). Nonetheless most of us see the importance and appropriateness of punishing and rewarding very young children for their actions because it is essential for moral training, to help them *become* moral agents who are responsible for their actions.

We see this in at least some circumstances with punishment of adults as well, particularly where legal punishment is concerned. For example, suppose a person, Nester, breaks the law by driving 10 miles over the speed limit on a stretch of highway. But suppose Nester was only speeding because the speed limit for that stretch of highway, a stretch he frequently drives on, had been temporarily reduced due to construction. And suppose that Nester had missed the sign indicating the temporarily lowered speed limit only because the excessively bright headlights of a passing car momentarily obscured his vision. Here it seems that Nester had a reasonable belief that he was driving the speed limit, and the fact that his belief was false was due to a minor distraction that was not his fault. And so I think we can plausibly assert that Nester was not guilty of any moral failing in driving faster than the speed limit. Nonetheless, it seems also plausible to say that no injustice is done if a police officer pulls Nester over and punishes him with a speeding ticket. It's still the case that Nester broke the law, and as far as the law is concerned ignorance of the law is no excuse. The speed limit needs to be enforced, even if sometimes those who break it do so without any moral negligence. Thus Nester is an apt target for punishment, even if it is not clear that he is morally blameworthy.

This point suggests to me that systems of punishment and reward are not *merely* retributive; an important component of any system of punishments and rewards is the

consequences that they have. This is not to say that I am endorsing a fully consequentialist theory of punishment (or any theory of punishment in particular - that would be beyond the scope of this project). It seems to me that notions of responsibility and desert are closely connected with and ought to inform our views on when we should punish and reward (and also what sorts of punishments and rewards are appropriate). I am only denying that a full account of appropriate reward and punishment is *identical to* (or necessary or sufficient for) moral responsibility.

III. Role Responsibility

Another sense of responsibility we often speak of has to do with one's duties or obligations insofar one occupies a certain position or station or office. This kind of responsibility has been called "role responsibility" by H.L.A Hart.[14] For example, a lifeguard has a responsibility to pay attention to what is happening on the beach and in the water, to ensure that everyone in the area remains safe. The lifeguard is therefore said to be *responsible for* the lives of those on the beach. This sense of responsibility is related to moral responsibility, even if they aren't one and the same. One way in which this sense of responsibility is related to moral responsibility is that you will typically be held morally responsible if you fail to live up to your role responsibilities. If the lifeguard shirks his duty to flirt with a cute girl on the beach, and if a small child drowns while he was distracted, we would probably say that he is morally responsible for the death of the child.

However, this is clearly not always the case. Sometimes it is possible to shirk one's responsibilities in this sense without in any sense being morally guilty or

[14] Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law*. New York: Oxford University Press, 1968.

blameworthy. The above example with Nester may be one instance of this; insofar as he occupied the role of a driver, he had a responsibility to obey the posted speed limit. And though he failed to live up to this duty, he did so (as I suggested) in a way that does not render him morally faulty. Or to take another example, imagine a soldier in the Confederate South who has been charged with tracking down a capturing a runaway slave. And suppose that this soldier, once he has caught up with the slave, has a crisis of conscience and decides to let the slave escape. This soldier, in virtue of his station, had a responsibility to capture the slave and return him to the South. Yet it seems clear that he had no *moral* responsibility to do so; quite to the contrary, it seems clear that his decision to allow the slave to escape was morally praiseworthy. While the roles we occupy do *often* create moral responsibilities, they don't always.

IV. Take Charge Responsibility

Another sense of responsibility that has been identified in recent work by Bruce Waller is "take-charge responsibility".[15] Take-charge responsibility is similar to role responsibility in the sense that it is a responsibility that one has in virtue of occupying a certain sort of position, but it is broader because it applies to the vast majority of people. It is the responsibility that we have in virtue of being people with some capacity to engage in reasoning and control our lives; it is the responsibility to take charge of our plans, projects, values, characters, etc. Just as the lifeguard has, in virtue of his position of lifeguard, some responsibility for the lives on the beach, so too do we all, in virtue of our status as reasoning agents, have some responsibility for our own lives.

A number of philosophers, both compatibilist and libertarian, have drawn a close

[15] See Waller, *Against Moral Responsibility*.

connection between this sort of "taken" responsibility and moral responsibility. On the libertarian side, Robert Kane acknowledges the objection that undetermined choices would be in an important sense arbitrary, since "the agents cannot in principle have sufficient or conclusive reasons for making one option and one set of reasons prevail over the other".[16] Kane notes that such an agent might, in spite of the fact that he lacked sufficient reasons for the action, note that he had *good* reasons choosing as he did and "stand by and take responsibility for"[17] the choice. On the compatibilist side, Frankfurt says something similar: "To the extent that a person identifies himself with the springs of his actions, he takes responsibility for them and acquires moral responsibility for them".[18] And along the same lines, Daniel Dennett argues that excuses like "it just didn't occur to me what harm I was doing" or "it was an accident" can and should be circumvented by the taking of responsibility. As he says, "healthy self-controllers shun this path. They *take* responsibility for what might be, very likely is, just an 'accident', just one of those things".[19] Similarly, Fischer writes, "One has control of one's behavior at least in part in virtue of having *taken control* of the mechanisms that produce it. One takes control by *taking responsibility."* [20]

It's clear that taking responsibility for ones actions and choices (as well as for the sort of person that one is) is extremely important, and there seems to be something very

---

[16] Kane in John Martin Fischer et al., *Four Views on Free Will*. Malden, MA: Blackwell Pub., 2007, 41.
[17] Kane in John Martin Fischer et al., *Four Views on Free Will*, 42.
[18] Harry Frankfurt, "Three Concepts of Free Action," *Aristotelian Society Proceedings Supplementary* 49, 122.
[19] Daniel C. Dennett, *Elbow Room: The Varieties of Free Will worth Wanting* (Cambridge, MA: MIT Press, 1984), 143.
[20] John Martin Fischer, *My Way: Essays on Moral Responsibility* (New York: Oxford University Press, 2006), 224.

attractive about drawing a connection between our acceptance of responsibility and moral responsibility. However, more has to be said. As tempting as it might be to do so, there are some problems with *equating* moral responsibility with responsibility that is taken, or with assuming that the two are coextensive. As Waller points out, we can easily construct a number of cases in which we can agree that someone has taken responsibility and yet it is still the case that there exists a clear dispute about whether that person is morally responsible.

To illustrate, consider a case of coercive manipulation (cases like these will be discussed extensively, especially in the next chapter). Imagine that a woman named Riley is at the beach, where she spots a young child drowning. She considers wading in to save the child (which she can do easily with no risk to herself), but instead decides to leave the beach and go to the movies. Unbeknownst to Riley, earlier that day a wicked hypnotist named Jesse had covertly hypnotized her and instilled in her an irresistible desire to go to the movies whenever she sees a child drowning. Riley is completely unaware of Jesse's hypnotic intervention. Later, when asked whether she is responsible for the death of the child that drowned, Riley says "Yes. I was in a position to save the child, and I chose not to, so I must accept responsibility for her death." Riley has certainly "taken" responsibility for her choice, but is she morally responsible? The overwhelming intuition is that she is *not* morally responsible because of the direct manipulation by Jesse, or that her responsibility is at least diminished. At the very least, it is clear that her responsibility is open to dispute; pointing to the mere fact that Riley "takes" responsibility for her actions is not by itself sufficient settle the question of whether she is morally responsible.

Drawing from considerations such as these, Waller concludes that while a person can indeed take responsibility in the take-charge sense, one cannot "take" moral responsibility. As he says, "because there are cases in which take-charge responsibility is clear and moral responsibility problematic, it is obvious that they are distinct and that establishing take-charge responsibility does not establish moral responsibility".[21] I agree with Waller that the two concepts are distinct, but I think nonetheless that they are closely connected. It seems plausible to me to say that take-charge responsibility *can* ground attributions of moral responsibility, given that some other important conditions are met. Waller is right that this claim requires further justification, which will be forthcoming.

V. Moral Responsibility

What, then, is moral responsibility? In my view, moral responsibility is best understood in terms of reactive attitudes. The underpinnings of this way of thinking about moral responsibility can be traced as far back as Aristotle, who discusses moral responsibility in terms of the appropriateness of reacting to agents with praise or blame for their actions. This way of thinking about moral responsibility gets developed in Peter Strawson's well-known 1962 essay, 'Freedom and Resentment'. According to Strawson, what we do when we hold someone morally responsible for an action is express an attitude towards them that is derived from our relationship to them and regard for them as person. Such attitudes include gratitude, love, forgiveness, resentment, anger, forgiveness, indignation, etc. In Strawson's view, the purpose of these attitudes (and thus the purpose holding someone morally responsible) is to express "how much we actually mind, how much it matters to us, whether the actions of other people - and particularly

[21] Waller, *Against Moral Responsibility*, 108.

some other people - reflect attitudes of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other"[22]. In short, our practice of holding others morally responsible for their actions is grounded in our connections to people, our interests in how they act, and what their actions reveal about the kind of people they are.

A number of other thinkers have followed Strawson in endorsing and developing this conception of moral responsibility. Fischer and Ravizza explicitly endorse Strawson's view of moral responsibility. As they say, "someone is a morally responsible *agent* insofar as he is an appropriate candidate for at least some of the reactive attitudes on the basis of at least some of his behavior (or perhaps his character)."[23] Similarly, R. Jay Wallace develops an account of moral responsibility that draws on Strawson's views. He writes: "On P.F. Strawson's view, emotions such as guilt, resentment, and indignation - what Strawson calls the reactive attitudes - provide the key to understanding moral responsibility and its conditions. I intend to develop this idea by working out an account of the stance of *holding* someone responsible, in terms of the reactive emotions."[24] Derk Pereboom, though he ultimately denies that we are ever morally responsible, also draws on Strawson's conception of responsibility. He agrees that some reactive attitudes, like indignation and moral resentment, are threatened by the loss of moral responsibility - and then goes on to argue that other important reactive attitudes, or at least elements of them,

---

[22] P. F. Strawson, *Freedom and Resentment, and Other Essays* ([London]: Methuen [distributed in the USA by Harper & Row, Barnes & Noble Import Division, 1974), 5.
[23] John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998), 6.
[24] Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994, 6.

that are not connected with moral responsibility survive, and that we can make due without the ones that are involved in moral responsibility.[25]

There are some apparent ambiguities in Strawson's original account, or at least some debates about how his claims about moral responsibility should be interpreted. For instance, from his original account one might draw the conclusion that what it means to hold someone morally responsible for an action is to feel a particular emotion. It seems clear that construing moral responsibility in this way would be a mistake. As Nomy Arpaly puts it, "If we simply follow the moral emotions to figure out who is blameworthy and who is not, we are bound to be misguided."[26] There are several reasons why this is the case. Which emotions a person happen to feel towards another in a given situation can obviously be influenced by a wide range of factors that have nothing to do with the moral blameworthiness or praiseworthiness of the target of the emotions. A number of factors might prevent me from feeling reactive emotions. Perhaps I have a particularly calm temperament, or perhaps I am physically exhausted, and so I fail to feel any sort of resentment or anger towards a friend who has betrayed me. Or perhaps my love for my brother mutes the annoyance and disappointment I might otherwise feel when he has failed to live up to his obligations. Or perhaps I might fail to feel gratitude towards a friend who has made a great sacrifice for me simply because I am selfish or absent minded. And similarly, I might for various reasons happen to feel reactive emotions in situations where it is inappropriate. For example, if I am a bad mood, then I might continue to feel resentment towards someone who has bumped into me even after I learn

---

[25] See Pereboom, *Living Without Free Will*, 199-207.
[26] Nomy Arpaly, *Merit, Meaning, and Human Bondage: An Essay on Free Will* (Princeton, NJ: Princeton University Press, 2006), 28.

that it was a complete accident. Or my affection for a friend may cause me to feel undue gratitude towards her when she has really done very little for me. And so on.

Considerations such as these make it clear that there is something more to holding someone responsible for an action than simply feeling a reactive emotion. Nevertheless, we may still say that moral responsibility is in an important sense grounded in such emotions. We should not conclude, as Arpaly puts it, that "reactive attitudes are readily felt signs of what is morally significant, but they should not be treated as what constitutes that significance."[27] On the contrary, the reactive attitudes seem to be a key element of the meaning of what it is to say that someone is appropriately held to be morally responsible.

To illustrate this, let us focus on one aspect of moral responsibility - blame. On any plausible account of moral responsibility, one implication of the claim that someone is a morally responsible agent is that she can be blamed for her misdeeds. "Blame" here is ambiguous however; we need to draw a distinction between *judging blameworthy* and *blaming* (this distinction is discussed by D. Justin Coates and Neal Tognazzini[28]). In blaming an agent, I adopt a certain attitude towards her; her moral transgression provokes an emotional reaction. As Wallace puts it, "To count as blaming a person, you have to be exercised by what they have done, and to be exercised in the relevant way is just to be subject to one of the reactive sentiments."[29] In other words, if I feel nothing at all towards your moral transgression - or if I have the wrong sort of emotional reaction to it, if I

---

[27] Arpaly, *Merit, Meaning, and Human Bondage*, 31.
[28] D. Justin Coates and Neal A. Tognazzini. "The Nature and Ethics of Blame." *Philosophy Compass* 7, no. 3 (March 2012): 197-207.
[29] Wallace, R. Jay., Rahul Kumar, and Samuel Richard. Freeman. *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. New York: Oxford University Press, 2011, 358.

praise you or admire you for it - then I cannot be said to blame you for it. By contrast, I can *judge* you blameworthy for a moral transgression without having any particular emotional reaction to it. In judging you blameworthy, all I do is come to the conclusion that a reaction of blame - with its requisite emotional component - would be warranted or appropriate.

With this distinction in mind, we can make better sense of the Strawsonian view that moral responsibility is grounded in the reactive emotions, without committing ourselves to the mistaken view that moral evaluations are nothing but emotional reactions. On the account that I (and others) favor, to say that someone is morally responsible is to judge them worthy of praise or blame - in other words, to judge that it would be appropriate to respond to his or her actions with some of the reactive attitudes (as stated in the Fischer and Ravizza quote earlier). This allows us to say that someone is morally responsible for an action without being committed to having any particular emotional reactions toward him. At the same time, it allows us to explain an important sense in which moral responsibility is grounded in the reactive emotions - moral responsibility is defined in terms of when those emotions are fitting or appropriate.

VI. Strengths of the Strawsonian Account

This approach to moral responsibility is unlikely to satisfy everyone. Some will insist on a fully metaphysical treatment of moral responsibility. As noted before, some will insist on tying moral responsibility in a very deep way with rewards and punishments. As Galen Strawson puts it, “true moral responsibility is responsibility of such a kind that, if we have it, then it *makes sense*, at least, to suppose that it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss

in heaven."[30] For those who are inclined to such approaches, an account of moral responsibility grounded in the reactive attitudes might seem in some sense to be shallow or lacking. Perhaps this is why compatibilist theories of freedom and responsibility (which often assume something like this account of responsibility) are dismissed as changing the subject,[31] or "redefining 'free will' to mean something else",[32] or engaging in a "quagmire of evasion".[33]

It is probably true that no one account of the phrase "moral responsibility" could do justice to everything that people have meant by it. Still, I think that the Strawsonian account of moral responsibility is a deep and robust one, and that it captures the core aspect of what is ordinarily meant by moral responsibility. This account of moral responsibility is not shallow, nor is it a dodge or part of an effort to redefine notions central to the debate over free will and determinism. On the contrary, I think that grounding our understanding of moral responsibility in the reactive attitudes, in the way described above, best enables us to explain some of its central features.

For one, this understanding of moral responsibility helps to explain why moral responsibility seems to matter so much to us. This point is developed nicely in Peter Strawson's original essay. When we understand moral responsibility as grounded in reactive attitudes like gratitude, love, forgiveness, resentment, and indignation, we can immediately see why moral responsibility is so important to us. Strawson talks of the

---

[30] Galen Strawson, "The Impossibility of Moral Responsibility," *Philosophical Studies* 75, no. 1-2 (1994): 9.
[31] See Harris, *Free Will*.
[32] Jerry A. Coyne, "Why You Don't Really Have Free Will," USATODAY.COM, January 1, 2012.
[33] James, William. *The Will to Believe and Other Essays in Popular Philosophy, and Human Immortality.* [New York]: Dover Publications, 1956, 149.

"very great importance that we attach to the attitudes and intentions towards us of other human beings".[34] The wills of others, as reflected in their words and deeds towards us, matter to us greatly. The reactive attitudes are in part expressions of this importance, and they themselves are deeply important to us. If we lose moral responsibility - if we conclude, as moral responsibility abolitionists like Pereboom and Waller do, that judgments of praiseworthiness and blameworthiness are always unwarranted - then we lose something of deep significance to us.

Of course, moral responsibility abolitionists like Pereboom and Waller want to deny that we actually lose anything of great significance if we give up moral responsibility. They argue that we can give up our practices of holding people morally responsible, and we can give up the reactive attitudes that go along with those practices, without losing anything of great significance or importance to our lives. I will, in due course (especially in the final chapter), examine these arguments and explain in detail why I think that we do in fact lose a great deal if it turns out that our judgments of praiseworthiness and blameworthiness are never warranted. At this point, it is sufficient for my purposes to note that even if critics like Waller and Pereboom are ultimately right, it is certainly not immediately *obvious* that they are right, and the Strawsonian account of moral responsibility at least explains the prima facie appearance that moral responsibility is integral to a full and meaningful life.

A second feature of moral responsibility that is best explained by understanding it as grounded in the reactive attitudes is that this account allows us to see why the debate over determinism and moral responsibility is so challenging. This advantage is made

---

[34] Strawson, *Freedom and Resentment and Other Essays*, 6.

most apparent when we contrast this account of responsibility with some of the others that have been mentioned. For example, consider the various accounts of moral responsibility that have been developed in terms of something like 'attributability'. Nomy Arpaly, for example, argues that whether a person is worthy of praise or blame is simply a question of whether their actions can be attributed to a good or ill will on the part of that person.[35] On her view, the central question is not whether a person has self-control, or is the ultimate author or source of that will, or other similar questions that many incompatibilists and compatibilists have tended to focus on. The question is just what a person's actions reveal about her will. In a similar vein, Michael Zimmerman describes moral responsibility in terms of what we can attribute to that person, "such that there are 'credits' or 'debits' in one's 'personal ledger', so that one is worthy of being judged to have such credits or debits".[36] The problem with this kind of approach is that it makes it somewhat mysterious just why the problem of moral responsibility and determinism has been such a persistent and vexing one. To put it another way, it is less than clear why the truth of determinism would pose a deep threat to moral responsibility if moral responsibility really just boiled down to a question of what we can attribute to a person.

One might try to argue (as some do) that if determinism is true then none of our traits or actions can really be attributed to us because they can ultimately be traced back to earlier external causes that precede us. However this does not seem like a very convincing reply. There is a clear common sense way in which we can attribute effects to specific salient causes, even if we know that the cause we attribute the effects to is itself

---

[35] See Arpaly, *Merit, Meaning, and Human Bondage*.
[36] Michael J. Zimmerman, *An Essay on Moral Responsibility* (Totowa, NJ: Rowman & Littlefield, 1988), 7-8.

an effect of earlier causes. For example, if a bolt of lighting strikes a barn and causes a fire, we have no trouble attributing the fire to the bolt of lightning. We have no trouble with this in spite of the fact that we are perfectly aware that the bolt of lightning itself was caused (by an electrical buildup due to atmospheric conditions), and knowing that this cause itself had causes - even if we assume that the chain of atmospheric causes and effects is all perfectly deterministic. By the same token, if Derek kicks a puppy for the thrill of hearing it yelp, there is no difficulty in attributing that action to Derek's ill will. This is true even if we know that Derek's ill will is the deterministic result of myriad earlier causes (genetics, an abusive upbringing, etc). The general point is that the existence of a chain of deterministic causes does not undermine picking out specific events in the chain and attributing later effects to them. When we seek causal explanations for events, practical considerations are relevant; some causes are more relevant to explanations than others, we are not required to focus exclusively on the earliest event in a causal chain. If you asked me why the barn was on fire, and I cited the fact that a butterfly flapped its wings long ago, I would have failed to give you the causal explanation you were looking for (even if the wing flapping is a necessary part of the causal chain that led to the fire).

Now it might be objected here that when our purpose is specifically *moral* responsibility, then other considerations apply - it must be the case that in some sense Derek, for example, is the "ultimate source" of his will, rather than his will being the result any deterministic antecedent causes. This is a more plausible consideration, and it is one I will address in detail, especially in the fourth chapter. But at this point we seem to have gone beyond an account of responsibility in terms of attributability (for again, it

seems clear that we can attribute Derek's actions to him whether or not he is the ultimate source of his will). In short, an account of moral responsibility in terms of attributability *alone* does not capture what is challenging about reconciling moral responsibility with determinism.

A similar point holds for accounts of moral responsibility that tie it closely with (or equate it with) punishment and reward, especially accounts that focus on the beneficial effects of punishment and reward (and/or the negative effects of doing away with punishment and reward). As mentioned before, a number of different philosophers writing on moral responsibility have made appeals along these lines. Here are some others. Moritz Schlick (claiming to draw from the work of David Hume) writes, "the question regarding moral responsibility is the question: Who, in a given case, is to be punished?"[37] Walter Stace also casually equates moral responsibility with punishment, writing "Thus we see that moral responsibility is not only consistent with determinism, but requires it. The assumption on which punishment is based is that human behavior is causally determined."[38] And similarly, Paul Gomberg argues along pragmatic lines, saying "from a practical point of view it may make good sense to hold the agent responsible for an action that he does because of something about himself (for example, callousness to the feelings of others) even though he is not responsible for that something about himself that is responsible for his acting as he did. By holding him responsible and acting accordingly we may cause him to shed an undesirable trait, and this is useful

---

[37] Moritz Schlick, *Problems of Ethics.* (New York: Prentice Hall, 1939), 152.
[38] Stace, "The Problem of Free Will", in Feinberg, Joel, and Russ Shafer-Landau. *Reason and Responsibility*. Belmont, CA: Wadsworth, 2013, 418.

regardless of whether the trait is of his making."[39] And likewise, Daniel Dennett defends the practices related to moral responsibility on practical grounds, writing: "Instead of investigating, endlessly, in an attempt to *discover* whether or not a particular trait is of someone's making - instead of trying to assay exactly to what degree a particular self is self-made - we simply *hold* people responsible for their conduct (within limits we take care not to examine too closely). And we are rewarded for adopting this strategy by the higher proportion of 'responsible' behavior we thereby inculcate."[40]

Even some critics and skeptics of free will also tend to think of moral responsibility as closely connected to our practices punishment and reward, and critique it on the basis of its effects. Noted free will skeptic Sam Harris discusses moral responsibility in these terms in places, writing that "responsibility is a social construct"[41] and arguing that we can continue to hold people responsible, blaming and punishing, provided it is true that these practices have beneficial effects.[42] Similarly Jerry Coyne, another noted free will skeptic, closely connects moral responsibility with punishment and reward. He writes: "But the most important issue is that of moral responsibility. If we can't really choose how we behave, how can we judge people as moral or immoral? Why punish criminals or reward do-gooders? Why hold *anyone* responsible for their actions if their actions aren't freely chosen?"[43] Coyne then goes on to explain how our practice of punishing criminals can still be justified on the basis of its practical effects, with the strong implication being that we should still hold people responsible even if they are

---

[39] Paul Gomberg, "Free Will as Ultimate Responsibility," *American Philosophical Quarterly* 15, no. 3 (July 1978): 208.
[40] Dennett, *Elbow Room: The Varieties of Free Will worth Wanting*, 164.
[41] Sam Harris, *The Moral Landscape* (London: Bantam, 2010), 210.
[42] Ibid., 142-148.
[43] Jerry A. Coyne, "Why You Don't Really Have Free Will"

unfree (while noting that "revenge" or "retribution" would be unjustified).[44] And as mentioned briefly before, Bruce Waller also draws a close connection in many places between moral responsibility and punishment; a substantial portion of his case against moral responsibility draws on the alleged negative effects that a system of punishment like ours has.[45]

A clear problem with thinking about moral responsibility this way - in terms of our practices of punishment and reward and the effects of those practices - is similar to the problem I discussed for models of responsibility based on attribution; it fails to capture or explain exactly what is so vexing about whether moral responsibility can be reconciled with causal determinism. If one takes this view of moral responsibility, then it is unclear (whether one is a skeptic or a defender of moral responsibility) why anyone would take it to be threatened by causal determinism. Whether the thesis of causal determinism is true or false has no clear bearing on the question of whether our current practices of punishment and reward (or something like them) are ultimately beneficial or harmful. That is a question to be settled by the social sciences, not by metaphysics.

VII. The Role of Metaphysics

Peter Strawson concludes that theoretical considerations, like the truth or falsity of causal determinism, are completely irrelevant to the question of whether our practices of holding people morally responsible for their actions are justified. Strawson argues instead that questions about the justifiability of judging people praiseworthy and blameworthy - and the reactive attitudes involved in those judgments - can only be

---

[44] I should also note that in other places Harris and Coyne write as if we should give up the notion of moral responsibility; their use of the concept is not always consistent.
[45] See especially Waller, *Against Moral Responsibility*, Chapter 8.

evaluated by principles *internal* to our practices. So if, for example, I have a reaction of indignation towards you for cheating at a game of poker, whether my reaction is appropriate will be determined by facts about the rules of the game, and our mutual attitudes about the rules of the game and agreed upon expectations about respecting the rules of the game, the history of our friendship, etc. External metaphysical considerations like the truth or falsity of causal determinism are, in Strawson's view, completely irrelevant.

In general, I am sympathetic with Peter Strawson's claim that the justifiability of our practices of judging people praiseworthy and blameworthy for their actions should be evaluated by principles internal to those practices. But I don't think it follows that metaphysical considerations are irrelevant. It seems clear that in at least some instances, metaphysical considerations *are* internal to our practices - that under ordinary circumstances, we intuitively recognize that theoretical considerations are relevant to evaluating the appropriateness of our reactive attitudes. As Galen Strawson puts it, "the roots of the incompatibilist intuition lie deep in the very reactive attitudes that are invoked to undercut it."[46] Wallace argues similarly in his exploration of what he calls a *generalization strategy*, in which one argues from widely recognized excuses or exemptions from moral consideration - many of which involve metaphysical considerations - to the conclusion that there is an incompatibilist requirement on moral responsibility.[47]

---

[46] Galen Strawson, *Freedom and Belief* (Oxford [Oxfordshire: Clarendon Press, 1986), 88.
[47] R. Jay. Wallace, *Responsibility and the Moral Sentiments*, 114-117.

Pereboom offers an example that supports the relevance of metaphysical considerations: "For example, some sexist and racist attitudes could be undermined by the following reflection: There is no difference across race and gender in capacities for theoretical and practical reasoning, for creative achievement, and for developing good human relationships. This reflection could and should radically alter human attitudes and practices, even if they are deeply rooted and longstanding."[48] Another example of this sort, discussed by Gary Watson, is of a man named Robert Harris who brutally murdered two teenage boys in 1978. Watson says, "we respond to his heartlessness and viciousness with moral outrage and loathing"[49] - in other words, we judge him an apt target for blame, and we blame him. But when we begin to learn in gruesome detail about the horrible abuse that Harris suffered as a child, this new knowledge "gives pause to the reactive attitudes."[50] We may not abandon our reactive attitudes of indignation completely, but it seems likely that they will at least be softened. And furthermore, our intuitive *judgments* of blameworthiness will likely be altered - we will judge that he is less deserving of blame that we did before. And it is at least plausible to say, as Pereboom puts it, that "our attitude of indignation is mitigated by our coming to believe that there were factors beyond his control that causally determined certain aspects of character to be as they were."[51]

So I do think it is pretty clear that metaphysical considerations do play a role in our ordinary practices of praising and blaming people, and of judging them worthy of

---

[48] Pereboom, *Living Without Free Will*, 98-99.
[49] Gary Watson, *Agency and Answerability: Selected Essays* (Oxford: Clarendon Press, 2004), 238.
[50] Ibid., 242.
[51] Pereboom, *Living Without Free Will*, 95-96.

praise and blame. The question remains, however, exactly which considerations play a role, and *how much* of a role do they play? And especially - does causal determinism play the role that incompatibilists think it does? On careful reflection, should we conclude that causal determinism would preclude the appropriateness of *all* judgments of praiseworthiness or blameworthiness? Examining that question will be the bulk of the work of the remainder of the dissertation. In the next sections, I will sketch a general framework for approaching this sort of question, drawing heavily on the work of some others. In the remaining chapters, I will use this framework to examine a series of arguments for incompatibilism and against compatibilism, arguing ultimately that they are all inadequate.

VIII. Standpoints

A number of philosophers emphasize a contrast between distinct "standpoints" in their discussion and analysis of the problem of free will and moral responsibility. Peter Strawson's approach is a prime example. Strawson distinguishes between two sorts of attitudes we may take towards other people - what he calls the *objective attitude* and what he calls *participant reactive attitudes*. If we take the objective attitude towards someone who has harmed us, then we regard him as an object thing to be dealt with - a thing that may be treated, or avoided, or removed from society, etc. In Strawson's view, we adopt the objective attitude towards others only when certain "exempting conditions" hold (like "he wasn't himself", or "he was delusional", or "she's only a kid"). By contrast, participant reactive attitudes are those attitudes we take towards others when we engage in personal relationships with them. As he puts it, they are "attitudes belonging to

involvement or participation with others in inter-personal human relationships"[52] - the sorts of attitudes involved in praising and blaming, as discussed in the previous section.

In Strawson's view, the key to resolving the free will problem is to understand these two distinct attitudes or standpoints, and the roles that the play in our relationships with and evaluations of others. He diagnoses both compatibilists (what he calls optimists) and libertarians and skeptics (what he calls pessimists) in terms of the distinction between these two standpoints. Optimists, Strawson says, make the mistake of ignoring the central role of the reactive attitudes in our lives, instead trying to justify our traditional moral practices in terms of their benefits or social utility (a number of traditional compatibilists do seem to argue in this way, as discussed in the previous section). Another way one might put Strawson's point is to say that optimists believe they can give a full account of and justification of moral responsibility in terms of the objective standpoint. Pessimists on the other hand do recognize that there is something deep and essential missing in the optimist's account of moral responsibility. However in Strawson's account, they misconstrue the nature of the relationship between the two standpoints, believing that theoretical considerations could and should (if determinism is true) lead us to adopt a purely objective attitude towards others.

There are a number of other philosophers who rely on a similar sort of distinction in explaining and diagnosing the problem of free will and moral responsibility. This general approach gets its most famous formulation in Immanuel Kant's distinction between practical reason and theoretical reason. While some (like Strawson, as noted above) use this kind of distinction between standpoints to defend a compatibilist view,

---

[52] Strawson, P. F. *Freedom and Resentment, and Other Essays*, 10.

it's not clear that Kant himself is pushing a compatibilist line here - in fact most people tend to characterize him as defending a libertarian view. For Kant, the distinction between the theoretical and the practical seem to have been not merely conceptual but also metaphysical, a distinction between the world of appearances and things in themselves. And Kant is famously and harshly critical of compatibilist approaches, referring to them as a "wretched subterfuge". For my purposes here I will leave aside the question of what Kant's own ultimate view was.

A number of other philosophers since Kant have followed in this tradition of evaluating the free will problem in terms of contrasting standpoints - Donald Davidson,[53] Thomas Nagel[54] (the contrast between subjective and objective viewpoints), and Daniel Dennett[55] (contrasting the "intentional stance" with other stances) are notable examples. Some draw compatibilist lessons from this sort of analysis, but others (like Nagel) push towards more skeptical conclusions. And some, like Kant (at least according to standard interpretations) and also Robert Kane[56] (at least to some extent, this will be developed later) draw libertarian conclusions.

Hilary Bok[57] has advanced a standpoint compatibilist account of this sort that I find particularly attractive, at least in some of its elements. In the next sections, I wish to discuss her ideas in a bit of detail, and show how they can be used to develop a

---

[53] See Davidson, Donald. *Essays on Actions and Events*. Oxford: Clarendon Press, 1980. (especially the essay "Freedom to Act")
[54] See Nagel, *The View From Nowhere*.
[55] See for example Dennett, Daniel Clement. *Freedom Evolves*. New York: Viking, 2003.
[56] See for example Kane, "Libertarianism", in *Four Views on Free Will*, 2007.
[57] Hilary Bok, *Freedom and Responsibility* (Princeton, NJ: Princeton University Press, 1998).

framework for examining various arguments related to the problem of free will and moral responsibility.

IX. Bok's Standpoint Approach

In Kantian fashion, Bok's standpoint approach relies on a distinction between theoretical reasoning and practical reasoning. For my purposes here there are a few points about the conceptual distinction between the theoretical and the practical. The first point is that the theoretical and practical standpoints have very different aims and employ very different concepts. The purpose of practical reasoning is to provide an answer to the question "What to do?" by analyzing and weighing reasons for different courses of action, so that a decision can be reached that can be regarded as justifiable. The purpose of theoretical reason, by contrast, is purely descriptive - to provide as complete and accurate a description as possible of the world, and of the objects therein and the causal relations that hold between them.

Engaging in practical reasoning gives us reason to employ concepts of norms and justification and value. The claims that practical reasoning arrives at can be particular (about what an agent has reason to do in a single instance) or general (about what an agent has reason to do in a type of situation, or generally). Practical reason draws on the data of theoretical reasoning, but theoretical claims alone cannot provide reasons or justifications. It is only in light of certain ends, purposes, or goals that a theoretical claim can serve as a justification for a certain course of action. And although practical claims often presuppose certain theoretical claims, the truth or falsity of those claims do not conflict with reasoning that led to those claims; rather, they simply show that one of the premises used in the reasoning was false, and so the reasoning was unsound. To use an

example of Bok's, my reasoning that I need to leave by 3pm in order to get to the doctor presupposes something about what time it actually is. If I learn that it's already 4pm, then I will no longer have reason to leave – not because there was an error in my practical norms or goals or reasons, but because my reasoning had a faulty premise.[58]

The key point about the conceptual distinction between practical and theoretical reason that Bok attempts to develop, and that I would like to developed further, is that reasoning from the practical standpoint gives us reason to employ a substantial concept of responsibility. We can begin to see this by noticing that engaging in practical reasoning involves examining our values, standards, reasons and goals to decide which our available alternatives is the best. And insofar as a person has values, standards, reasons, and goals, she has reason to value character traits that help us act in accordance with our values and reasons and standards and achieve our goals. So if furthering my career is a goal of mine, then I have reason to value traits like diligence and persistence that help in advancing one's career. Of course, interest in a trait can be defeated if it conflicts with other standards or values or goals that a person holds. Ruthlessness might also be a valuable character trait for a person interested in advancing her career, but insofar as she also holds herself to moral standards, she will lack interest in developing that particular character trait.

Bok argues that insofar as the practical standpoint gives us reason to be concerned with our character traits, it also gives us reason to examine our past actions, to explore whether it was the action I ought to have performed, given my standards and what I knew (or ought to have known) at the time. In particular, I will have interest in exploring

---

[58] Bok, *Freedom and Responsibility,* 69.

whether a past action reveals any flaws in the quality of my will, any character traits that work against the standards I hold myself to. Insofar as my past actions are determined by my practical reasoning, I have reason to regard my past actions not merely as events that befell me, but as things that I did, either in accordance or not in accordance with my principles. And if my past actions are not in accordance with my principles, then I have reason to try to correct the character defects that led to them, else they will continue to lead to actions that go against my principles. In short, from the practical standpoint I have reason to distinguish between those actions of mine that reflect my will and character and those behaviors that do not, and to hold myself *accountable* for those which do. Though in some sense I am of course causally responsible for all of my actions and their results, it is only in the former case that the practical standpoint gives me any reason to critically evaluate and to lament (or to celebrate) the things that I have done.

There are a couple of kinds of actions that might fall in the category of actions that do not reflect my will or my character, and it is in exploring these that I can begin to show how this sort of standpoint approach to compatibilism connects with and can supplement other approaches. The first kind includes actions that may be done as the result of an irresistible psychological impulse or addiction. For instance, consider the true kleptomaniac – one who is subject to impulses to steal such that it is literally impossible for her to resist, no matter how much she tries and no matter how much she recognizes and is moved by the reasons (moral and otherwise) for not stealing. From the practical standpoint the kleptomaniac lacks reason to criticize herself for past acts of stealing simply because the action did not flow from and is not responsive to her practical reasoning – to her evaluation of what, with respect to her goals, values, and principles,

she ought to do. And the same seems to hold for the addict. Insofar as the man who is addicted to alcohol is truly unable to resist the desire for another drink, he lacks (from the practical standpoint) reason to hold himself responsible for drinking – again precisely because his evaluations of whether or not he ought to drink (given his values, principles, etc) are causally inefficacious. A number of philosophers have emphasized the central importance of responsiveness to reasons to moral responsibility (see Fischer and Ravizza,[59] and Arpaly,[60] for prominent examples). So I think it is interesting and instructive that instances in which the practical standpoint gives one reason to hold oneself responsible are just those situations in which one is actually responsive to reasons, in the way suggested by such accounts. Thus, it seems to me that an account of responsibility grounded in the nature of practical reason can mesh with and bolster the reasons responsiveness account of moral responsibility by providing a way of explaining its intuitive plausibility.

In addition to those cases in which one is unresponsive to reason, there is another sort of case in which the practical standpoint would fail to give us any reason to hold ourselves responsible. This is the sort of case in which our actions are not the result of our own deliberations and evaluations of what we ought to do, but rather are the direct result of someone *else's* evaluation of what we ought to do. In other words the practical standpoint gives us no reason to hold ourselves responsible in just those sorts of manipulation cases that are so problematic for compatibilist accounts in the way I described earlier. I think this insight will help point the way towards an answer to

---

[59] Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998.
[60] Arpaly, *Merit, Meaning, and Human Bondage*.

"manipulation" objections to compatibilism - a point I will return to in much more detail in the fourth chapter.

V. Is this Moral Responsibility?

I have suggested so far that there is a sense of responsibility that is justified from the standpoint of practical reason - that insofar as we engage in practical reason, it makes sense and is appropriate for us to assess the extent to which we are acting as we believe (all things considered) that we should, to hold ourselves accountable for how well we succeed at or fail at living up to our standards. A question one might have at this point is whether the concept of responsibility justified from the standpoint of practical reason is really a robust concept of genuine *moral* responsibility. It seems that so far, it falls somewhat short of the mark. For illustration, I will bring in the story of the Scorpion and the Frog – a well-known fable sometimes attributed to Aesop. The story is simple enough. A scorpion is trying to figure out a way across a river. Being a scorpion, she is of course unable to swim across. So she asks a frog to carry her across the river. The frog is at first hesitant, worried that the scorpion will sting him. But the scorpion reminds the frog that if she were to sting the frog, the frog would sink and they would both drown. Reassured, the frog begins to carry the scorpion across the river. Mid-way across the river, the scorpion stings the frog. As they both sink, the frog asks the scorpion why she doomed them both by stinging him. The scorpion replies simply, "I am a scorpion, it is my nature".

Is the scorpion morally responsible for stinging and killing the frog? One response (for those who emphasize reasons responsiveness) might be that the scorpion is not morally responsible because she is not appropriately responsive to reasons. After all there

is clearly a very strong reason for the scorpion not to sting the frog, namely that it will result in her own demise. So it might seem that the scorpion is compulsively acting on her desire to sting in a non-reasons responsive way. But suppose this isn't the case – suppose it is not a compulsion, but simply the scorpion's deepest desire. The scorpion understands that stinging the frog will have fatal consequences but doesn't mind; she judges that the pleasure of stinging the frog is more important than even survival. In this case the scorpion is responsive to reasons, it just so happens that the joy of stinging is a reason for her that takes precedence over survival. So if we understand the example in this way, the scorpion can plausibly be said to satisfy something like Fischer and Ravizza's reasons-responsiveness condition for moral responsibility. Similarly, we may even say that the scorpion's higher order desires mesh with her lower order desires – then she satisfies the conditions of a classical hierarchical account of freedom.[61] We could build in other conditions as well. On the face of it, the scorpion seems able to satisfy any of a variety of compatibilist accounts of what it takes to be morally responsible.

The problem is that there is a strong prima facie reason for us to say that the scorpion is *not* morally responsible for stinging the frog. As the scorpion says, it is simply her nature to sting the frog. In other words, it is in virtue of her being a scorpion that she has the character that he does. And of course the scorpion can't help being a scorpion, and thus she can't really help having the character that he does, and thus it seems she cannot really be morally responsible for the actions that flow from it. It might be argued that this is a defect of the various compatibilist accounts, and that it illustrates an advantage of libertarianism. A libertarian could argue that if the character of the

---

[61] See for example Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68, no. 1 (January 1971): 5-20.

scorpion is fully determined by the kind of creature that it is, then she is not responsible for the character (since the scorpion cannot control what it is) and the actions that flow from it.

There are a couple of ways that compatibilists might respond. One way is to simply bite the bullet and say that the scorpion is indeed morally responsible. And at first glance this may not seem wholly implausible. It might be argued that the fact that the scorpion has such a malicious, monstrous character and endorses it is enough to say that the scorpion is morally responsible, even though the scorpion ultimately cannot control what her nature is. But it's not clear to me that this is actually the right sort of response; I think that a more nuanced response to the example is required.

The other option for compatibilists is to try and provide a principled explanation for why the scorpion is in fact not morally responsible. I think that considering the requirements of practical reason can be helpful here, and in this way we can also answer the question of whether the concept of responsibility that you get from the practical standpoint alone is in fact a full concept of moral responsibility. So now let us consider things from the standpoint of the scorpion, as an agent engaged in practical reasoning. As an agent deliberating about what sorts of actions are best for it to do, the scorpion has reason to employ some sort of a concept of responsibility that distinguishes between the actions that flow from her will and those that don't, as has been argued above. And so if the scorpion failed to act on certain sorts of goals that she valued, she would have reason to criticize herself for the failure, to hold herself accountable for it. But does the scorpion have any reason to regard herself as *morally* responsible for stinging the frog - in other words, to regard herself as an apt target for moral praise and blame? If we assume that the

scorpion is indifferent to moral ends then it seems that she does ***not*** have any such reason. And so it seems that the requirements of practical reason alone do not ground a full concept of ***moral*** responsibility, even if they do ground *some* sort of concept of responsibility. Pereboom makes a similar point about Bok's account, arguing that the sense of responsibility that Bok defends is an important one, but not full sense of moral responsibility.[62]

What this suggests to me is that engaging in practical reason does not, *by itself*, provide one with reason to employ a full concept of moral responsibility. Instead, it is only if one has moral goals or concerns or ends, in addition to being a practical agent, that one will have reason to employ such a full and robust concept. Interestingly, I think it is instructive here to consider to what Bok herself has to say about the moral emotion of guilt. Bok argues that the emotion of guilt has often been misconstrued, regarding guilt as a self-directed punishment. This has led many thinkers (e.g. Sigmund Freud[63], William James[64], and Friedrich Nietzsche) to dismiss or condemn guilt. Some, like Freud, might regard it as performing an important social function, but at best it is a necessary evil needed to constrain our impulses. At worst, guilt might be seen as pointless and narcissistic - "a kind of self-flagellation that harms us without in any obvious way benefiting those we have wronged."[65] This way of construing guilt encourages the view that there is something disingenuous about the emotion, that it doesn't track anything real. As Nietzsche put it, "Although the most acute judges of the witches and even the

---

[62] Pereboom, *Living Without Free Will*, xxi-xxii.
[63] Freud, Sigmund. *Civilization and Its Discontents*. New York: W.W. Norton, 1962.
[64] James, William. *The Varieties of Religious Experience: A Study in Human Nature*. New York: Modern Library, 1936.
[65] Bok, *Freedom and Responsibility,* 168.

witches themselves, were convinced of the guilt of witchery, the guilt was nevertheless non-existent. It is thus with all guilt."[66]

Bok argues that this is a misleading way to view guilt. Guilt, appropriately felt, is not like self-inflicted punishment. Instead, "it is like the relation between the recognition that one's relationship with someone one truly loves has collapsed and the pain of heartbreak. Heartbreak is not a pain one inflicts on oneself as punishment for the loss of love; it is not something we undergo because we deserve it; nor need we develop our susceptibility to it to meet some antecedent conception of the kinds of responses that situations of this kind warrant. Rather, supposing that what one has lost was in fact love, and that one has indeed lost it, one can fail to suffer only if one walls oneself off from these facts, either by blinding oneself to their existence or by denying their importance. Pain is not only an appropriate response to the collapse of love but the only response that accurately registers what has happened."[67]

I find this comparison between the feeling of guilt and the feeling of a loss of love to be insightful, and I think it helps to point the way to an understanding of moral responsibility. The basic idea is that the feeling of guilt is simply the pain that we experience when we recognize that we have violated our moral standards. The pain of guilt is a fitting and appropriate response to violating one's moral standards because those standards define what matters to us - they are what we think of as valuable, or how we think is the good or right or appropriate way to live. To know that we have violated our values - to have failed regard others appropriately or to live in the ways we think a

---

[66] Walter Kaufmann and Friedrich Wilhelm Nietzsche, *The Portable Nietzsche* (New York: Penguin Books, 1954), 96-97.
[67] Bok, *Reason and Responsibility*, 169.

person ought to live - *should* be an emotionally painful experience. On this account, the appropriateness of guilt has nothing to do with its instrumental value (whether it is instrumentally valuable at all is a completely separate question); rather, the claim is simply that the kind of pain that characterizes guilt is the fitting reaction to the recognition that one has failed to live as one recognizes that one ought to.

I think that this way of characterizing guilt points the way to answering some of the questions that have been raised so far in this chapter. For starters, it gives us the resources to explain why it is that the scorpion in the story is not morally blameworthy. She is not blameworthy because she lacks the moral concerns (and even the capacity for the moral concerns) that would make a feeling of guilt fitting or appropriate; from her standpoint, she has not violated any important values. While the scorpion in the story is indisputably a wicked, monstrous creature, she is not a participant, so to speak, in the moral community. Since the scorpion is not a participant, since she utterly lacks the values or standards that would make the moral emotion of guilt appropriate (or even possible) for her, regarding her as an apt target for moral blame strikes me as inappropriate as well. So on this point, I depart a bit from Bok's view. Unlike Bok, I do not think that the requirements of practical reason alone can ground moral responsibility. To get to *moral* responsibility, one's practical reasoning must be infused with an understanding of and genuine concern for moral ends and moral goals, including an appropriate susceptibility to the relevant reactive moral emotions (like guilt), before judgments of praiseworthiness and blameworthiness can be said to be appropriate.

This way of viewing guilt also helps to defuse some of the possible objections to Bok's account of moral responsibility as grounded in practical reason. Some might argue

that Bok's account of kind of concern that practical reason gives us for the quality of our will is too forward looking, or too instrumental. But when we recognize that the nature of moral emotions like guilt have nothing to do with instrumental value – that the appropriateness of guilt lies simply in the nature of the recognition that one has violated ones moral values – then this sort of concern is diminished.

This way of characterizing guilt also helps to show how Bok's account of moral responsibility can be connected, in a mutually supportive way, with Peter Strawson's account. In Bok's view, engaging in practical reason gives us reason to employ certain concepts of responsibility. And as I argued, drawing on Bok's account of guilt – when we have the right moral ends, this responsibility will take on a particular emotional character. In the case of guilt, we will feel the pain of violating our moral standards, and we will (if we properly understand guilt) recognize the appropriateness of that pain. We should then see guilt as an *apt reaction* to our moral wrongdoing, and from there, I think it is a fairly short step to recognizing that we are apt targets of reactive attitudes (and judgments about the appropriateness of those attitudes) from others. This also helps point to a response to Waller's claim that we cannot merely "take" moral responsibility. With the understanding of responsibility in terms of the appropriateness of praise and blame, we *can* (in a sense) take moral responsibility – we take it under those conditions when, as agents engaged in practical reasoning with moral values and moral emotions, we are capable of recognizing that personal reactive emotions like guilt (or also shame, pride, etc) would be appropriate reactions to our own success or failure at living up to our moral standards.

VI. Conclusion

In this chapter, I have explored some accounts of responsibility, and have defended a broadly Strawsonian account of moral responsibility as the one that is most fitting with our ordinary conceptions. I have drawn on standpoint approaches to free will and responsible, especially Hilary Bok's account of freedom and responsibility from the standpoint of practical reason, to explain how our engaging in practical reason with moral ends can help to ground claims of moral responsibility in the Strawsonian sense. Admittedly, this account is somewhat sketchy so far. At this point, it is a broad framework for viewing the problem of free will and moral responsibility. What remains to be seen is whether the standpoint of practical reason with moral ends gives us reason to accept exempting conditions for the appropriateness of reactive attitudes because of the kinds of metaphysical concerns that typically move incompatibilists (concerns like access to alternate possibilities, or ultimate sourcehood of our characters and wills, and others). These aspects of my account will be fleshed out in greater detail in the coming chapters, beginning with a look at "leeway" incompatibilist concerns in the next chapter.

Chapter 3 – Leeway Incompatibilism

In the history of the debate over the compatibility of freedom and responsibility with causal determinism, two main strategies of advancing the incompatibilist viewpoint have emerged. This chapter will focus on one of those strategies – what has come to be called "leeway incompatibilism". According to the leeway incompatibilist, the primary explanation for why causal determinism rules out genuine freedom and moral responsibility is because it eliminates our ability to act otherwise than we actually do. In other words, for the leeway incompatibilist, the problem with causal determinism is that it means that each person's life is on a fixed, preset path with no branches – no leeway - as opposed to the "garden of forking paths" that genuine freedom and responsibility require. Another way we might put the claim is this: in order to be someone who can truly be deserving of praise or blame for an action, it has to have been possible that you could have done something to avoid that praise or blame.

I. Conditional Analysis

Compatibilists have traditionally responded to leeway arguments in one of two different ways. One classical compatibilist response (suggested at least as far back as David Hume)[68] and expounded upon by countless later compatibilists like A.J. Ayer (see Ayer, 1954) and R.E. Hobart (see Hobart, 1934) is to argue for a *conditional analysis* of the ability to do otherwise than one actually does. According to the conditional analysis, claims about the ability to do otherwise are claims about what would happen given certain counterfactual conditions. The rough idea is that we should analyze statements of the form "John could have done X" as "John would have done X if John had wanted (or

[68] See for example section 8.1 of David Hume, *An Enquiry Concerning Human Understanding, 1748*. Oxford: Oxford UP, 1999.

willed, desired, preferred, decided - the analyses vary) to do X". This kind of analysis has considerable prima facie plausibility. It draws a distinction between those things that we have the power to do if we so choose to do them and those things which we would fail to do even if we attempted to do so. And since counterfactual claims about what a person would have done given alternate conditions don't conflict with deterministic claims about what will happen given the actual conditions, this analysis (if correct) provides a way to secure the compatibility of the "Garden of Forking Paths" model of freedom with causal determinism.

The problem is that the conditional analysis of 'could have done otherwise' espoused by classical compatibilists runs against some powerful counterexamples. The analysis gets the wrong results in a range of cases, suggesting that agents have the ability to do otherwise in situations where it is clear that they do not (Peter van Inwagen argues for this claim at length).[69] Here is one classic example of that sort from Keith Lehrer.

> Suppose that I am offered a bowl of candy and in the bowl are small round red sugar balls. I do not choose to take one of the red sugar balls because I have a pathological aversion to such candy. (Perhaps they remind me of drops of blood and …) It is logically consistent to suppose that if I had chosen to take the red sugar ball, I would have taken one, but, not so choosing, I am utterly unable to touch one.[70]

In this example, Lehrer satisfies the conditional analysis of freedom regarding his choice to refuse the red candy. We may suppose that if Lehrer had chosen (or preferred, decided, wished, etc) to take one of the candies, then he would have. So the conditional analysis yields the result that Lehrer has the ability to take a candy. Yet it seems clear that (given

---

[69] Peter van Inwagen, *An Essay on Free Will* (Oxford [Oxfordshire: Clarendon Press, 1983), 55-105.
[70] Keith Lehrer, "Cans without Ifs," *Analysis* 29 (1968): 32.

that his pathological aversion to the red candies makes it impossible to so choose) he is in fact *unable* to take one. In general, it seems that the conditional analysis of freedom, at least as classically formulated, isn't capable of handling restrictions to freedom that come from *within*.

In light of these sorts of counterexamples, some philosophers have sought to refine the analysis. One way we might refine the analysis is to tack on extra conditions. So we might end up with an analysis like "An agent can do X just if (i) If the agent were to choose to do X, the agent would do X, and (ii) The agent is not hindered by a phobia or other psychological disorder, and not subject to hypnosis or other manipulation, and not under the influence of any drugs, and so on". The problems with such an analysis should be obvious. The analysis is incomplete - how many things get included under the "and so on"? And what principled connection is there between the list of conditions that remove an agent's ability to do otherwise? Does causal determinism get included on the list? The compatibilist cannot simply exclude determinism from the list of defeating conditions without justification, or else he completely begs the question against the incompatibilist.

Other philosophers have come up with more interesting and insightful proposals for explaining how the ability to do otherwise than one actually does might be analyzed in a way that is consistent with determinism (see for example Donald Davidson,[71] Chrisopher Peacocke,[72] and Keith Lehrer[73]). I will not explore in detail here whether any of them ultimately work. Instead, I will focus most of my attention on the second

---

[71] Donald Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), 63-83.

[72] Christopher Peacocke, Being Known (Oxford: Clarendon Press, 1999).

[73] Keith Lehrer. "'Can' in Theory and Practice: A Possible Worlds Analysis." In *Action Theory*, edited by Myles Brand and Douglas Walton, 241-70. Dordrecht: D. Reidel, 1976.

compatibilist strategy, in particular the work of Harry Frankfurt. Instead of trying to reconcile the ability to do otherwise with causal determinism, Frankfurt rejected the idea that the ability to do otherwise than one actually does is actually necessary condition for moral responsibility. In what follows, I will explain and defend Frankfurt's strategy.

II. Frankfurt and the Principle of Alternate Possibilities

To begin, let's explore the central principle of the leeway incompatibilism in more detail – the principle that says that in order for a person to be responsible for an action, that person must have been able to do otherwise, to have refrained from doing the action in question. This Principle, dubbed "The Principle of Alternate Possibilities" (or PAP) by Harry Frankfurt,[74] has obvious intuitive appeal. Here is an example for illustration. Imagine a girl named Riley is walking down a beach and notices a small child drowning in the ocean. She is in a position to see that she could save the child easily and with no risk to herself. But instead, Riley chooses to walk away from the beach. Quite predictably, the child drowns a few minutes later. Intuitively it seems obvious that Riley is morally blameworthy for doing nothing to save the child. But now suppose we add a wrinkle to the story - suppose Riley is actually crippled by a debilitating fear of water (descriptively, we can call this case "Phobia"). In fact, we learn that the only reason that Riley was even on the beach was as an initial attempt to desensitize herself to being in the presence of large bodies of water like the ocean. Riley very much wanted to save the child, but the very thought of approaching the ocean filled her with debilitating terror, and she was overwhelmingly compelled to turn and walk away from the source of her terror.

---

[74] Harry Frankfurt. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66, no. 23 (1969): 829-39.

With this new information, it seems clear that Riley is *not* morally responsible for choice to turn and walk away (or at the very least, her responsibility is greatly diminished). And the proponent of PAP has a ready way to explain this new intuition. What we learned about Riley is that her possible courses of actions were restricted. Due to her aquaphobia, she was not able to do otherwise than turn and walk away, and restriction on her abilities is what makes it so that she cannot be held responsible for that failure.

However, as Frankfurt pointed out, the plausibility of PAP is undermined when we consider another sort of case. Frankfurt-style cases (or Frankfurt examples or scenarios) are stories in which there are clearly no alternatives for the agent in question, and yet it remains intuitively obvious that the agent is responsible for his or her action. To demonstrate, let's consider a revised version of Phobia. Suppose that Riley has no fear of water at all. Suppose that she considers saving the drowning child, but realizes that doing so might make her a few minutes late for a movie that she'd like to see that afternoon (call this version of the example "Movie"). And so, Riley walks away and leaves the child to drown. Unbeknownst to Riley, a wicked neuroscientist named Jesse has implanted a microchip in Riley's head that carefully monitors her brain activity. If Riley had begun to show even the slightest inclination that she might decide to save the child, the chip would have taken over Riley's brain and forced her to walk away. But as it happens, Riley acted purely of her own volition, and the chip remained causally inert. In this case, it seems that it wasn't possible for Riley to do otherwise than walk away from the beach. Nonetheless, the intuitive judgment that she is responsible for her action remains. It would be very implausible to claim that the presence of a chip that plays

absolutely no causal or explanatory role whatsoever in how Riley *actually* acts could do anything to diminish her moral responsibility. The lesson of Frankfurt examples is that what matters for attributions of moral responsibility is what happens in the actual sequence of action; the presence or absence of alternate possibilities is irrelevant.

III. A Dilemma for Frankfurt Examples

Naturally, Frankfurt's rejection of PAP hasn't convinced everyone. There are a few objections to Frankfurt examples that I would like to consider in some detail. One objection can be put in the form of a dilemma. Riley's decision to walk away and let the child drown was either causally determined (by earlier mental states, environmental factors, etc.), or it wasn't. On the first horn of the dilemma – if the decision was not causally determined, then there is no way for Jesse's chip to predict with certainty what Riley will do. If the decision was causally determined then the chip can make its prediction, but then there is no reason for the libertarian to accept the conclusion that Riley is morally responsible. We have simply begged the question against the libertarian by asserting that Riley is morally responsible in causally deterministic scenario. So either way, it is argued, Frankfurt cases fail to give us any reason to reject PAP.[75]

There are at least a couple of different compatibilist responses to this objection. One is to question whether it's really true that it is impossible to predict free undetermined actions – I'll return to this point later. But first, let us consider a response from Fischer. Fischer concedes that any Frankfurt scenario is likely to have *some* sort of alternate possibility built into it. But even granting this, he thinks we can construct successful Frankfurt scenarios. As he points out, if PAP is to have any plausibility at all,

[75] For one example of this sort of objection, see Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), 142-143.

the "alternate possibility" needs to be something robust - it needs to be something that can be reasonably asserted as a ground for moral responsibility. But as Fischer demonstrates, it is easy to construct scenarios in which the precursor is something minor and involuntary, not nearly robust enough to plausibly ground attributions of moral responsibility. Returning to the Movie example - suppose that a precursor to Riley deciding to leave and allow the child to drown is that she would blush slightly. This precursor is reliable – it always precedes Riley deciding to dismiss the obvious need of the child for the sake of making it to the movie on time. This can be the sign that Jesse watches for; if Riley does not display the physiological signs of a slight blush by a certain time, then the microchip will take over.

In this case, we may suppose that there is indeed an alternate possibility – Riley might have failed to blush. But it is a very minor, involuntary alternate possibility, a mere "flicker of freedom"[76] (to borrow Fischer's phrasing) - hardly the sort of thing that could be a plausible candidate for grounding moral responsibility. As Pereboom suggests, what is lacking with an alternate possibility like not blushing is that an agent could not avoid moral responsibility for his or her action *by* securing that alternate possibility.[77] In other words, if Riley had managed not to blush, she would not *in virtue of not blushing* avoid moral responsibility for letting the child drown; the failure to blush would not explain her lack of moral responsibility.

---

[76] For an extended discussion of this argument, see John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control* (Cambridge, MA: Blackwell, 1994), 134-147.
[77] Derk Pereboom, *Living without Free Will* (Cambridge, U.K.: Cambridge University Press, 2001), 7-8.

IV. A Dilemma for Precursors

In response to this, some incompatibilists have pressed a new version of the dilemma mentioned earlier. Suppose we construct a case just as Fischer suggests, where the precursor is something minor and involuntary like a blush. Then we are left with only two possibilities - either the precursor is causally sufficient to ensure the desired action, or else it isn't. If the precursor IS causally sufficient to ensure the manipulator's desired action (if blushing means that Riley will definitely leave the child to drown), then the ensuing action is causally determined by the precursor, and it is then open to the libertarian to insist that Riley is not morally responsible. In other words, as before, it begs the question against the libertarian to assert that an agent like Riley could be morally responsible for an action that is causally determined by an involuntary reaction like blushing. If, on the other hand, the precursor is NOT causally sufficient to ensure the desired action, then there may be room for the libertarian to argue that there IS a "robust" alternate possibility. That is, if the blushing is not enough to ensure that Riley in fact walks away and leaves the child to drown, then that means that it IS possible for her to save the child even if she blushes. If it remains possible for her to save the child, in spite of the occurrence of the precursor, then there is an option that remains open by which she could avoid responsibility for the death of the child - namely, choosing to try and save the child.[78]

This objection is more forceful, and fashioning an adequate response is challenging. Compatibilists have tried to respond in a few different ways. One strategy of

[78] For a development of this line of criticism, see David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *Philosophical Review* 104 (1995), 247-261.

response is to try and put indeterminism somewhere *before* the precursor, allowing the precursor to causally determine the ensuing action (for examples see Eleanore Stump[79] and Ishtiyaque Haji[80]). Another strategy involves constructing scenarios in which the precursor is *necessary* but not sufficient for the ensuing action, while also showing that robust alternate possibilities have been eliminated for the agent in question.[81] And a third strategy involves trying to construct Frankfurt scenarios which eliminate the need for a precursor altogether. While the first two strategies have much to be said for them – and may in fact be successful at dealing with this problem (in particular, I find Pereboom's strategy to be plausible), my main focus is going to be on defending a strategy of the third sort. I think that doing away with precursors altogether is the most promising way of building robust Frankfurt examples that can handle not only this objection, but also another powerful sort of objection that has been raised in recent work by Kadri Vihvelin (I will return to this last point in later sections of this chapter).

V. Blockage Cases

There are a couple of prominent examples of ways of developing Frankfurt examples without precursors. The strategy I wish to focus on is developed by David Hunt,[82] in his advancement of what have come to be called "blockage"[83] cases (for another way of developing Frankfurt examples without precursors, see Al Mele and

---

[79] Eleanore Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," in *Faith, Freedom, and Rationality: Philosophy of Religion Today*, ed. Jeff Jordan and Daniel Howard-Snyder (Lanham: Rowman & Littlefield, 1996), 73-88.
[80] Ishtiyaque Haji, *Moral Appraisability: Puzzles, Proposals, and Perplexities* (New York: Oxford University Press, 1998), 36.
[81] For an example of this sort, see Derk Pereboom, *Living without Free Will*, 18-28.
[82] David Hunt, "Moral Responsibility and Unavoidable Action," *Philosophical Studies* 97, no. 2 (1997): 195-227.
[83] This term appears in John Fischer, "Recent Work on Moral Responsibility," *Ethics* 110, no. 1 (October 1999): 114.

David Robb).[84] To illustrate Hunt's approach, I will construct an example that resembles his, using the cast of characters that I have been using so far. To begin, we imagine two different scenarios. **Scenario A** resembles the Movie case above; Riley is at the beach, she sees the child drowning, and she walks away and lets the child drown so that she can get to a movie on time. In this scenario, Riley possesses full, robust, undetermined, libertarian free will – however one may wish to cash that out. And so it seems clear that Riley is morally responsible for letting the child drown.

In **Scenario B**, we imagine a version of Riley that is *exactly* identical to the Riley of Scenario A, right down to the atom, with one exception: a nefarious neuroscientist named Jesse has blocked the neural pathways in Riley that would allow her to make any alternate choices. The causal history of Riley B's decision is *exactly* the same as the causal history of Riley A's; Jesse does nothing to causally determine Riley B's choice. The only difference is that for Riley B, neural blockages exist that make it impossible for her to do anything other than walk away from the beach. Is Riley B morally responsible for walking away and letting the child drown? Since her causal history is identical to Riley A's, it seems that we must say that she is morally responsible. Since Jesse did not determine her decision in any way – she did nothing but subtract alternate possibilities that play no role in the actual causal sequence of Riley B's decision and ensuing action – it is hard to see how Jesse's action did anything that could diminish Riley B's blameworthiness.

---

[84] Al Mele and David Robb, *Rescuing Frankfurt-Style Cases* 107, no. 1 (January 1998): 97-112.

However some critics of blockage cases have objected that they ultimately amount to deterministic scenarios, and thus there is no reason for an incompatibilist to accept that Riley B is morally responsible for her actions. One way to raise the deterministic worry is with a question – as Fischer puts it, can the neurons in Riley B's brain "bump up" against the neural blockage, or can't they?[85] If the neurons *can* bump up against the neural blockages, then there is room to argue that there are robust alternate possibilities that ground Riley B's moral responsibility. If they *cannot* bump up against the neural blockages, then it might seem that by putting the blockages in place, Jesse has in effect causally determined Riley's action.[86]

However it is not immediately clear why a defender of Frankfurt examples should accept the claim that blockage cases are (or amount to) cases of causal determinism. Granted, Riley B's action is inevitable – but it is not inevitable because of its causal history - her causal history is the same is Riley A's. In this sense, it is no different than other Frankfurt scenarios that seek to make a certain action inevitable without tampering with the actual causal history. Pereboom expresses this point lucidly, as follows:

> In response, one might point out that in the standard Frankfurt-style cases, the relevant action is inevitable, but the intuition that the agent is morally responsible for it depends on the fact that it does not have an actual causal history by means of which it is made inevitable. What makes the action inevitable is rather some fact about the situation that is not a feature of its actual causal history, and hence the action's being inevitable need not make it the case that it is causally determined. But then how is the blockage case different from the standard Frankfurt-style cases? After all, the blockage does not seem to affect the actual causal history of the action.[87]

---

[85] Fischer, "Recent Work on Moral Responsibility", 119.
[86] This view is argued for by Robert Kane, "Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson," *Philosophy and Phenomenological Research* 60, no. 1 (January 2000): 157-167.
[87] Pereboom, *Living Without Free Will*, 17.

For these reasons, our intuitions about blockage cases should be no different from our intuitions about other Frankfurt scenarios. All that blockage cases do is make it more clear that access to alternate possibilities has been cut off, and they do so without relying on any problematic precursors.

Insofar as there is a legitimate lingering deterministic worry, I think it has to do with the fact Jesse has to be able to *predict* what Riley is going to do in order to be able to block all of (and only) the neural pathways that in fact would not have been activated. This is similar to the worry that was raised in the first dilemma I presented for Frankfurt scenarios. As Kane writes in this telling passage:

> there *are* no alternative possibilities left to the agent; every one is blocked except the agent's choosing A at t. But now we seem to have determinism pure and simple. By implanting the mechanism in this fashion, a controller would have predetermined exactly what the agent would do (and when); and, as a consequence, the controller, not the agent, would be ultimately responsible for the outcome. Blockage by a controller that rules out all relevant alternative possibilities is simply predestination; and on my view at least, predestination runs afoul of ultimate responsibility.[88]

I find this passage telling because of the way in which Kane raises the issue of predestination, and especially the way he casually equates it with determinism. In my view, this is the only significant difference between blockage cases and other sorts of Frankfurt examples, like the ones that rely on precursors. I think it is clearly a mistake to say that *Jesse* predestines what Riley B will ultimately decide to do. If Jesse is able to ascertain in advance what Riley B will choose, and then if it is predestined at all, it is predestined before Jesse does anything to block off any alternate possibilities. So the question is whether Jesse's access to future facts about what Riley B will choose to do is

[88] Kane, "Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson", 162.

itself a threat to freedom, whether it amounts to determinism. I will address this question in the following sections.

VI. Foreknowledge, Future Facts, and Moral Responsibility

A preliminary point I would like to make is that ordinary foreknowledge of a person's choices is no more threatening to their freedom or moral responsibility than the mere existence of future facts about that person's choices. To put the point another way - what is (at least apparently) threatening about the claim that someone knows in advance what I will choose is that such knowledge presupposes that there is a fact about what I will choose, and that this fact obtains before I make more choice. The mere addition of an agent who has somehow managed to ascertain this fact[89] presents no extra threat.

Let us consider another example for illustration. Suppose at some point T2 in the future, I will be making a choice between going out to get Thai food or Indian food for dinner. And I suppose it is true now, at T1, that I will in fact choose to get Thai food at T2 (call this example Thai). And suppose further that there is some other agent, John, who knows that I will get Thai food at T2. The apparent threat to my freedom can be expressed by the fact the following two propositions seem to be incompatible:

P1: It is true at T1 that I will, at a later time T2, choose Thai food.

P2: It is true at T2 that I was free to choose Indian food over Thai.

Further propositions about who *knew* about the outcome of my future choice don't seem to add anything to the threat. Consider the following additional proposition:

---

[89] The exact details of how he might ascertain this fact aren't too important, so long as the agent's beliefs are not *necessarily* true, as in the case of essential divine omniscience. For a discussion of the issues particular to a divine being that possesses essential omniscience (as God is typically defined), see John Martin Fischer, "Freedom and Foreknowledge," *The Philosophical Review* 92, no. 1 (January 1983): 67-79.

P3: John knows at T1 that I will, at a later time T2, choose Thai food.

If P3 is incompatible with P2, it is purely in virtue of the fact that it entails P1; it is hard to see any other reason that P3 could be a threat to P2.

The question I want to consider now is - should a libertarian reject the compatibility of P1 and P2? To put it another way - should acceptance of the claim that a free choice (like my choice to get Thai food) implies the falsity of causal determinism imply that one ought also accept that my free choice for Thai food at T2 implies that no statements like P1 are true? Or to put it yet another way, bringing the point closer to the discussion at hand - does PAP imply (or should one who accepts PAP also be committed to) the claim that there are no facts about what we will do in the future? I think the answer to this last question is clearly no. It would be unreasonable to absolve someone of moral responsibility for an action merely because it had been a fact that the action would be committed at some point in the past before it occurred. I think that consideration of a few points will make this obvious.

First, the existence of future facts like that expressed in P1 implies nothing about the truth of *causal* determinism. It might be that some people have the intuition that the existence of future facts would restrict freedom because they implicitly assume that the existence of future facts would mean that the future has been causally determined by the present; if so, then the intuition is based on a mistaken assumption. There could facts about future events – the events could in some sense already be "out there" – and those facts would be causally undetermined by, even completely causally unconnected to, anything in the present (the current state of the universe, the laws of physics, etc). There

is no inconsistency in the claim that there are facts about the future in a causally indeterministic world.[90]

Second, to claim that the mere existence of future facts deprives us of freedom or responsibility is to commit the error of logical fatalism. The idea of logical fatalism is an old one, dating back at least to Aristotle's famous sea battle problem. The problem can be summarized as follows. Suppose that there is going to be a sea battle tomorrow between the Greeks and the Trojans. The statement "The Greeks win the battle" is presently either true or false (by the law of excluded middle). Suppose that it is true. If it is true now that the Greeks will win the battle, then it is impossible that they lose. It is generally agreed that this reasoning commits a modal fallacy – that we shouldn't infer from the fact that the Greeks will win the battle tomorrow that it is *necessary* that they will win tomorrow. This kind of fallacy will be discussed in a bit more detail in later sections when we consider Kadri Vihvelin's argument against Frankfurt examples (in which she accuses their defenders of committing a very similar sort of fallacy).

Third, the claim that the mere existence of facts about how one will act in the future deprives one of freedom or moral responsibility seems unmotivated by any of the standard libertarian considerations or intuitions. The sorts of considerations that motivate libertarianism (and incompatibilism in general) have to do with being genuinely in control of oneself, not determined or constrained by outside causal factors that are

---

[90] For a related discussion, see Michael Levin, "Compatibilism and Special Relativity," *The Journal of Philosophy* 104, no. 9 (September 2007): 433-63. Levin explores (and rejects) arguments by philosophers who claim that the "tenseless" model of the universe implied by special relativity entails that we are unfree. Of particular relevance to the present discussion is that, as Levin notes, even those who draw fatalistic conclusions from special relativity acknowledge that it is consistent with causal indeterminism (see page 440).

beyond ones control. Libertarians are generally concerned with having complete control over one's own choices and actions. They want it to be the case that one is, in some sense, the *ultimate source* of one's own actions - and causal determinism seems to rule this out, because one's actions would ultimately be traceable, via a long chain of causes, to sources (earlier events and the laws of physics) outside the self (e.g. *source compatibilism*). Similarly, Libertarians typically want it to be the case that one's choices *unrestricted* - that outside factors do not place limitations on what one can choose, that one can always choose from some range of options (*leeway incompatibilism*). Again, causal determinism (it can at least be plausibly argued) rules this condition out, because a deterministic causal chain allows for only one possibly outcome for any decision or action. These libertarian conditions for freedom are at least prima facie plausible; it is not unreasonable to be moved by these sorts of considerations, at least initially.

The question, then, is whether the mere existence of future facts legitimately raises either of these sorts of worries. As far as I can see, it does not. Return to the earlier example of my future choice to go out to get Thai food for dinner. The truth of P1 in no way raises any source concerns when I actually make my choice at T2. There is no feasible sense in which the past fact about my present choice is the cause or source or explanation or reason for my action. It is perfectly consistent with the mere existence of a past fact about my present choice that I am a "prime mover unmoved",[91] that my action is a "self-willed" or a result of my "self-formed character",[92] that there is no aspect of my decision that can be traced in any sort of causal or explanatory way beyond my pure,

---

[91] Roderick Chisholm, "Human Freedom and the Self," in *Free Will*, ed. Derk Pereboom (Indianapolis: Hackett Publishing Company, 1997), 152.

[92] See Robert Kane's account of freedom in terms of self-forming actions in Kane, *The Significance of Free Will*, 1996.

unfettered, uninhibited will. In other words, there is no reason to suppose that the past fact about my future choice expressed in P1 is a fact about a completely free (in a totally libertarian sense) choice.

Similarly, I don't think that the truth of P1 raises any legitimate leeway concerns. If leeway concerns have any sort of force at all, it is because there is something apparently threatening to our freedom and responsibility about the idea that some sort of outside force could *constrain* our choices, limiting us to only one option. The driving force behind the initial intuitive plausibility of PAP is that it doesn't seem like we possibly could be held responsible for an action when some external force or cause (like the distant past plus the laws of physics) makes it so that we could not have possibility done anything else. But again, it is hard to see how the mere existence of facts about our future choices in any way places a outside constraint or limitation on what we may do. If I am a libertarian agent - if my actions have no external causes, if my choices are purely a product of my own will and nothing else, then it is hard to see how I am constrained. How can a mere fact about what I do of my own will itself be a constraint on my will? I think we can make the implausibility of this notion even more clear if we connect it to a case involving moral responsibility.

Consider a version of the earlier example where I have done something morally significant – where I have ditched my fiancé Bunny by choosing to go for Thai food rather than to the Indian restaurant I agreed to meet her at. She would undoubtedly be inclined to blame me for deciding to ditch her. Now suppose I offered the following sort of excuse: "I'm sorry I didn't meet you as planned, but I really had no choice. Getting Thai was the only thing I could do, it was impossible for me to go to the Indian restaurant

we agreed to meet at." This sounds like a reasonable excuse at first. Of course it would be fair for Bunny to inquire as to *exactly why* it was impossible for me to meet her at the Indian restaurant. She might wonder if I had been kidnapped, or forced at gunpoint, or if not that, then what other external force made it so that I could only go for Thai food? If I replied that there was no external force whatsoever - that I had been constrained by the existence of a fact about what I would choose to do of my own uncaused, purely self-determined will, one could hardly fault her if she continued to blame me for my choice. If causal determinism were true, and I could point to some past deterministic factor that made it so that I could do nothing other than choose Thai food, then at least I would have *some* sort of case to make in pleading that I don't deserve blame. But if I am libertarian free agent, then it is hard to see what possible moral excuse I have, whether facts about what we will do in the future presently exist or not.

A fourth point worth mentioning here is that if libertarians were to insist that the mere existence of facts about the future is incompatible with freedom, regardless of whether those facts are causally determined by anything prior, then their theory of freedom becomes even more hostage to contingent facts about physics and metaphysics than it already is. As Fischer has argued, one of the unattractive features of libertarianism is that it leaves moral responsibility to "hang on a thread",[93] in the sense that our status as agents who can be appropriately praised and blamed for our behavior would hinge on the deliverances of theoretical physicists, cosmologists, and metaphysicians. In other words, it seems quite peculiar to suggest that we can't know for sure how to appropriately regard and treat other people and ourselves until we settle arcane questions of microphysics and

[93] John Martin Fischer, *My Way: Essays on Moral Responsibility* (New York: Oxford University Press, 2006), 5-6.

cosmology. The libertarian who insists on the non-existence of facts about the future would just make this problem drastically worse – the thread by which our moral responsibility hangs would be made that much thinner. For now not only does our status as responsible agents hinge on what physics and metaphysics might one day tell us about causal determinism, it also hinges on what they might tell us about the existence of future facts. Indeed, if it is true (as many argue)[94] that Einstein's theory of special relativity entails the existence of facts about what will happen in the future, then the libertarian who insists on their non-existence may have already cut the thread.

VII. Practical Reason and Alternate Possibilities

At this point, it will be helpful to consider whether the existence of future facts presents any threat to the exercise of practical reason, or whether practical reason even requires any sort of robust notion of alternate possibilities at all. By considering this question, we can give further support to Frankfurt examples (including the blockage versions of the sort devised by Hunt), and provide further principled reason for abandoning the "alternate possibilities" requirement on moral responsibility.

A number of incompatibilists (and some compatibilists) have tried to motivate leeway requirements for freedom and moral responsibility by appeal to the requirements of practical reason. Such thinkers argue that the truth of determinism would in some way threaten or undermine the activity of practical reasoning - that is, the activity of deliberating about reasons for goals or desires or aims, and on that basis deciding which future courses of action to take, presupposes the existence of alternate possibilities. Initially at least, one can see the plausibility of this worry. But ultimately I will argue that

[94] See again Levin, "Compatibilism and Special Relativity", 2007.

this worry is misguided – whether one thinks that the restrictions on alternate possibilities comes from causal determinism, or from the existence of facts about what we will choose to do in the future. To see this, let us explore a bit more just what it is that some philosophers have found threatening about causal determinism to our ordinary activity of deliberating about what to do.

Many philosophers have argued that deliberation between options requires some sort of genuine belief in freedom, in the sense of having alternate possibilities. Some would argue that this is a belief in alternate possibilities that is compatible with determinism, but many others use this claim to support incompatibilism. The latter sort would argue from the claim that determinism entails that there is only one possible course of action that deliberation requires a belief in the falsity of determinism; they would argue when determinists deliberate (as of course they do), they are able to do so only because they hold inconsistent beliefs.

John Searle, for example, writes: "Consider any situation of rational decision making and acting and you will see that you have a sense of alternate possibilities open to you."[95] Richard Taylor argues for the impossibility of deliberation under deterministic assumptions, writing: "If one does not know what he is going to do, but knows that conditions already exist sufficient for his doing whatever he is going to do, then he cannot deliberate about what to do, even though he may not know what those conditions are."[96] Peter van Inwagen writes: "one cannot deliberate about whether to perform a

---

[95] John R. Searle, *Rationality in Action* (Cambridge, MA: MIT Press, 2001), 15.
[96] Richard Taylor, "Deliberation and Foreknowledge," *American Philosophical Quarterly* 1, no. 1 (1964): 76.

certain act unless one believes it is possible for one to perform it."[97] Ishtiyaque Haji argues similarly, saying: "lack of freedom to do otherwise subverts the truth of judgments of practical reason; thus, assuming that determinism precludes our having alternatives, determinism undercuts the truth of such judgments."[98] Hector-Neri Castaneda suggests that if the universe is such that we lack genuine alternate possibilities, then "the universe is ugly … we are thus condemned to presuppose a falsehood in order to do what we think practically."[99] Taking a similar but slightly softer view, Randolph Clarke claims that while it isn't *impossible* for one to deliberate in the face of the belief that only one alternative is really possible, "the presumption is practically inescapable on a consistent basis."[100] Thus he also agrees that if determinism is true, we are destined to be subject to an illusion at least quite often (maybe nearly always) when we deliberate.

Let's grant, just for the sake of argument, that the truth of the thesis of causal determinism entails that there is only one possible course of action for any agent in any given situation. Does that it mean it is impossible to deliberate (or least, impossible to do so authentically or consistently) if one believes in the truth of determinism? To put it another way, would deliberation then presuppose a belief in libertarian free will? Many philosophers (typically compatibilists, of course) have rejected this claim. Many of them endeavor to show, through the use of examples, that it is entirely possible to deliberate

---

[97] Inwagen Peter. Van, *An Essay on Free Will* (Oxford [Oxfordshire: Clarendon Press, 1983), 154.
[98] Ishtiyaque Haji, *Freedom and Value: Freedom's Influence on Welfare and Worldly Value* (Dordrecht: Springer, 2009), 119-120.
[99] Hector-Neri Castañeda, *Thinking and Doing: The Philosophical Foundations of Institutions* (Dordrecht: D. Reidel, 1975), 134-135.
[100] Randolph K. Clarke, *Libertarian Accounts of Free Will* (Oxford: Oxford University Press, 2003), 113.

(authentically, legitimately, consistently, rationally) even when one *clearly* recognizes that there is only one possible outcome. Here is one such example from Clarke:

> Imagine that Edna is trying to decide where to spend her vacation this year. She mentions this fact to her friend Ed, who, as it happens, is in possession of information that Edna does not yet have. Ed knows that Edna will soon learn that she can, with less expense than she had expected, visit her friend Eddy in Edinburgh. And given what Ed knows about Edna and her other options, he knows that after she learns of this opportunity, she will eventually decide to take it. However, Ed is a playful fellow, and he doesn't tell Edna all of this. He tells her only that he knows that she will eventually learn something that will persuade her to spend her vacation with Eddy in Edinburgh.[101]

Clarke argues that Edna might believe Ed (and that Ed may indeed be right), and that Edna can nonetheless still deliberate about where to take her vacation. But one might argue that what Edna would be doing in this case would actually count as practical reasoning. If she accepts that Ed is right about where she will wind up vacationing, then it could be argued that she is no longer truly deliberating about where to go, because she has already reached her conclusion about where she will go. As Neil Levy argues, her reasoning now seems more theoretical than practical; "she is seeking to discover what reasons there are in favor of her choosing Edinburgh."[102]

To avoid this sort of worry, the compatibilist needs is an example in which an agent doesn't know which choice he is going to make, where he nonetheless recognizes that only one of the options he is considering is actually available to him, and where he is engaged in genuine deliberation. Peter van Inwagen argues that such an example is impossible. He likens deliberation in the face of the knowledge that only one option is a genuine option to the following kind of example, asking anyone who wonders whether

---

[101] Randolph Clarke, "Deliberation and Beliefs About One's Abilities," *Pacific Philosophical Quarterly* 73 (1992): 108.
[102] Neil Levy, "Determinist Deliberations," *Dialectica* 60, no. 4 (2006): 456.

deliberation requires belief in alternate possibilities to "imagine that he is in a room with two doors and that he believes one of the doors to be unlocked and the other door to be locked and impassable, though he has no idea which is which; let him then attempt to imagine himself deliberating about which door to leave by."[103]

In van Inwagen's example, it seems clear that (rational, consistent, authentic) deliberation is impossible. In this example, you know that your deliberation will not be efficacious; only one door is open, and no matter how you deliberate, that is the door you will be going through. There is no other possibility. Deliberation in light of this fact is pointless; the only way you could begin to deliberate is to ignore the truth that only one possibility is truly open to you. But is this example really analogous to what it would be like to deliberate in the face of the recognition of the truth of causal determinism? It seems clear to me that it is not. In van Inwagen's case, we can't deliberate about which door to go through - but we can simply try the doors and find out which door is unlocked. This is radically unlike ordinary cases of deliberation between two options. Ordinarily when we deliberate between two options, we cannot simply try and find out which option is possible (for example, I cannot simply test to find out which option, Thai or Indian, is "unlocked") - I have to make a decision. This highlights the central disanology in van Inwagen's example - in his example, which door is unlocked is completely independent of the decision that I make. In ordinary cases of deliberation, the decision that I make IS the option that will be "unlocked".

Hilary Bok develops a different example to capture this feature of ordinary deliberation. She describes it as follows:

---

[103] van Inwagen, *An Essay on Free Will,* 154.

> Instead of imagining that I am in the situation van Inwagen describes, we should instead imagine that I am in a room with two doors, one of which is locked and one of which is unlocked; that I do not know which is which; but that I do know that the locks are set up in such a way that as soon as I choose to try to open one door, that door will unlock, and the door I have not chosen will lock. We should also imagine that these doors do not open onto the same hallway. Perhaps one opens onto a Tahitian beach and the other into downtown Manhattan. Would it be irrational to deliberate in this situation?[104]

This scenario is better - it more closely resembles what ordinary deliberation is like. In Bok's version of the scenario, which door winds up locked is now tied to my deliberation. If I go through one door, it will be the case that it is impossible for me to go through the other. Unlike van Inwagen's case - and like ordinary deliberation - I cannot simply try the doors to see which is unlocked, because trying one will mean that the other option is closed off to me. I am forced to deliberate and decide which door I most have reason to open.

However Bok's scenario also has a shortcoming. The way it is described, it is not entirely clear that there is only one door I can go through. Since one door only becomes locked *after* I attempt to open the other one, it seems that I genuinely could choose either door - neither option is closed off to me. At the moment of decision, either door could become the locked door or the unlocked door. Even if this is not how Bok intended the example to be read, the way she described the scenario at least leaves it open to this reading - and this may be what explains our intuition that we can genuinely deliberate about which door to open.

So what is needed is a scenario in which it is made much more explicit that there is truly only one option open to us at the time of deliberation, and in which it nonetheless remains very clear that genuine, authentic, rational deliberation remains possible. Neil

---

[104] Bok, *Freedom and Responsibility*, 111-12.

Levy seeks to develop just such a case, borrowing the idea of a predictor from classic formulations of Newcomb's problem. The scenario is as follows:

> Sally finds herself in a room with two door. She knows that one of the doors is locked and the other unlocked, but she doesn't know which is which. She also knows that the unlocked door is the door that the predictor (an alien super-scientist, let us suppose), who is able to predict the decisions of human beings with 100% accuracy, has predicted that Sally will choose.[105]

Levy stipulates that Sally believes - correctly, it seems - "that she is not metaphysically free to open either door, since one door is locked and will remain locked whatever she does."[106] It also seems that Levy has described a deterministic scenario (the alien super-scientist's perfect knowledge of the deterministic causes of human behavior seem to be behind his ability to predict human behavior with 100% accuracy) - so no indeterministic or incompatibilistic assumptions are being smuggled in. And yet nonetheless it is clear that it makes perfect sense for Sally to deliberate in this case, because she knows that whichever door she decides upon is the door that is unlocked.

I find Levy's version of the scenario convincing. It seems clear that Sally should regard herself as having only one option metaphysically open to her (we may even add to the story that she knows in some detail exactly how the alien, with its knowledge of human neurophysiology, etc., is able to make its deterministic predictions, to make this point more explicit), because one door has been locked before she ever started deliberating. And yet it also seems equally clear that she can deliberate about what she is going to do. All that seems necessary for genuine, rational deliberation is that her choice - via the alien scientist's prediction – is causally efficacious, that it was the factor (via the alien's reliable prediction) that determined which door would be unlocked.

---

[105] Levy, "Determinist Deliberations", 456.
[106] Ibid., 457.

Some incompatibilists may remain unconvinced. An incompatibilist might insist that insofar as Sally is engaging in deliberation, that she does not truly regard one of the doors as closed to her - that she must have inconsistent beliefs.[107] But it is not at all clear why we should suppose that this is the case. That is, there is no reason that Sally must believe that both doors are genuinely open to her in order to deliberate. All that matters for deliberation is that the door she decides that she has the most reason to open is the one that will be open. Levy's case has that key feature, while making it explicit that there is only one genuine option for Sally.

Perhaps an incompatibilist will be bothered by the apparent fact that in the described scenario Sally *could* deliberate, but she doesn't need to. In Levy's case, Sally can simply try the doors to learn which one is open. And so perhaps an incompatibilist would argue that in a scenario like the one Levy describes - one in which only one option is open to an agent - deliberation is pointless or superfluous, even if it is technically possible. Perhaps this sort of worry is why Bok designed her scenario so that the agent could not simply try the doors to discover which was open.

However I don't think that this sort of objection is particularly threatening to the claim that deliberation is compatible with the existence of only a single option. Even if it was the case that Sally could just "try" the doors and find out what the upshot of her deliberation would have been, this point does not generalize to deliberating in ordinary deterministic scenarios in which only one option is (arguably) open to an agent to perform. In ordinary causally deterministic scenarios, (like the one where I am trying to choose between Thai and Indian food) an agent cannot just "try" and find out where

---

[107] See E. J. Coffman and Ted A. Warfield, "Deliberation and Metaphysical Freedom," *Midwest Studies in Philosophy* 29, no. 1 (2005): 25-44.

practical reason would have led him. In ordinary scenarios agent cannot find out where practical reason will lead until he or she actually deliberates and makes a choice.

And in fact, this general point seems to apply to Sally as well. If Sally decides to just "try" one of the doors at random to find out which is locked and discover where practical deliberation would have led her, then the alien super scientist will have predicted that she would have done this, and will have left whichever door she first grabbed for at random unlocked. Thus, Sally will not actually have discovered anything about where practical reason would have led her by trying a door at random. If Sally actually wants to discover where practical reasoning will lead her, then she has to actually engage in practical reasoning - she has to deliberate about which door she has the most reason to open. This is true in spite of the fact that there is only actually one door that she can open. Thanks to the alien super scientist with perfect powers of prediction, it remains perfectly rational for Sally to deliberate about the outcome, even when the outcome has been determined in advance.

At this point we can tie the discussion with the "blockage" cases raised by Hunt. Suppose that instead of locking the door that it predicts Sally won't choose, the alien instead tampers with Sally's brain. Using its perfect knowledge of human neurophysiology, the alien predicts exactly which door Sally is going to choose. Then, it blocks off all of the alternate neural pathways that would have been activated had she been going to choose otherwise. The other door is locked, so to speak, in her mind; the blockages make it impossible for her to choose the other door. And suppose the alien informs Sally of all of this.

Can Sally rationally, consistently deliberate about what to do, knowing as she does that there is only one possible outcome to her deliberation? I say that she clearly can - because the blockages are sensitive to the upshots of her ordinary deliberative process. If she decides to give up and just open a door at random, then the alien will have predicted that, and that will be the only thing she could do. So there is no practical reason for her to engage in that course of action. She has every reason to figure out which door she ought to open, using just the sort of practical reasoning she would ordinarily use in that scenario. In this case it is vividly clear that there is only one thing she can do - it's hard to see how she could maintain inconsistent beliefs once she understands what the alien has done to her brain - and nonetheless it is perfectly obvious that she can, and should, still deliberate. So from the standpoint of practical reason, there is nothing problematic or threatening about Hunt's "blockage" version of the Frankfurt scenario. This provides stronger reason to think that Hunt's version of the scenario is successful; this kind of example provides strong reason to think that whatever moral responsibility might require, it is not connected to the existence of alternate possibilities.

VIII. Vihvelin's Challenge

Now I would like to turn the discussion to an entirely different sort of objection to Frankfurt examples developed by Kadri Vihvelin. She argues that Frankfurt stories rely on a modal fallacy, and that as a result, they fail to actually show that there are situations in which a person could not do otherwise while remaining morally responsible for his or her actions. To begin with, she argues that an important distinction has been missed in the traditional discussion of Frankfurt examples, and that as a result that Frankfurt examples have seemed to be more plausible than they should. The distinction is between

"conditional interveners" and "counterfactual inteveners". The former are explained as follows: "What makes someone a conditional intervener is the fact that his intervention is causally triggered by the beginnings of any action (overt or mental) contrary to the intervener's plan. If the subject begins to try or begins to do any undesired action, the intervener will prevent him from succeeding."[108]

By contrast, the counterfactual intervener does not respond to anything that the agent actually does. Instead:

> His intervention is causally triggered, not by the subject's trying or beginning to act contrary to the intervener's plan, but by some earlier event that is a reliable indicator of the fact that the subject will, in the absence of intervention, choose or act contrary to the intervener's wishes. The earlier event might be a blush, twitch, or other involuntary sign that occurs just before the subject begins to make an unwanted decision.[109]

Of the former - the conditional intervener - Vihvelin states that examples using such an intervener fail to show that PAP is false. This is because conditional interveners wait for the agent to do something, and insofar as they have do so, it is open for the incompatibilist to argue that this alternate something that the agent could do is a necessary condition for moral responsibility. Of the latter - the counterfactual intervener - Vihvelin again argues that such an intervener fails to eliminate alternate possibilities. She argues that people who claim that they do are guilty of making a logical mistake - an error in modal reasoning. She illustrates this with an example. I will present a slightly simplified version of the example (rather than quoting the entirety of her lengthy description).

---

[108] Kadri Vihvelin, "Freedom, Foreknowledge, and the Principle of Alternate Possibilities," *Canadian Journal of Philosophy* 30, no. 1 (March 2000): 9.
[109] Ibid.

Suppose that you and I make a bet on the outcome of a coin toss. I bet heads, and you bet tails. Unbeknownst to you, I have a nefarious confederate named Black who can always predict coin tosses accurately (Vihvelin doesn't explain how - perhaps he is psychic). Furthermore, Black has the ability to change the predicted outcome of the coin toss if he wants to. Black is, for whatever reason, interested in making sure that I win our bet. On the morning of the coin toss, Black peers into the future and sees that I will win - that the coin will come up heads all on its own, without his intervention. So he retires for the day, doing absolutely nothing to affect the outcome of the coin toss. Later that evening, we place our bets, toss the coin, and indeed - just as Black predicted I would - I win the bet.

Vihvelin claims that in this scenario, I have won the bet fair and square. The game would be rigged only if Black had intervened. Since Black didn't intervene - since he had retired for the day, and had nothing whatsoever to do with the coin toss - it would be completely illegitimate for you to complain that I do not deserve to win the bet. The bet remains fair because the coin, which had in no way been tampered with, still could have come up tails.[110] The fact that Black exists as a counterfactual intervener ensures that it wouldn't come up tails, but so long as the scenario is such that he doesn't actually intervene, he has done nothing to make it so that it couldn't come up tails. If we reason from the mere fact that the coin will come up heads that it must, that it was necessary that it come up heads, then we have simply made a mistake in modal reasoning - the same sort of mistake that is made by the logical fatalist.

---

[110] Of course, if you continued to play the game with me after learning all the facts, you would be crazy - but that is only because you have learned now that there are many days (those days in which Black does actually intervene) in which the coin toss is *not* fair.

Vihvelin then argues that the sort of mistake that you would be making if you claimed that the game was rigged - that the coin couldn't have come up tails - is exactly the same mistake that is made by defenders of Frankfurt scenarios claim that the counterfactual intervener makes it so that an agent could not have done otherwise than he actually does. For example, let us return to the "Movie" example that I discussed earlier, in which blushing is a precursor to Riley deciding to leave the beach and let the child drown. In this scenario, Jesse fits Vihvelin's definition of a counterfactual intervener. She watches for the sign that tells her whether she should intervene, and in the actual scenario, she never has to intervene. Because Riley exists as a counterfactual intervener it is ensured that Riley won't save the drowning child. But what Vihvelin is suggesting is that it would be a mistake to infer from that fact that Riley couldn't save the child - that she has been deprived of any ability. That is, in Vihvelin's view we should still say that Riley still *could* save the drowning child, even though we know she won't - and if we say that, then it remains open for the incompatibilist (or the classical compatibilist) to say that Riley's ability to do otherwise than leave the child to drown is a necessary condition for the intuitive judgment that she is morally blameworthy for that decision. In short, examples that make use of counterfactual interveners are not counterexamples to PAP.

This is a novel and powerful objection. If Vihvelin is right, then the path that Frankfurt set many compatibilists on in decades of dismissing PAP has been a massively misguided diversion. In what follows, I want to briefly explore a few ways that a Frankfurtian compatibilist can and should respond to Vihvelin's argument (there are undoubtedly others).

IX. Replying to Vihvelin

For starters, let's consider the coin-flipping scenario. Is Vihvelin right that the coin still could come up tails, even though it won't? It seems so. The mere fact that Black can reliably predict the results of the future coin flip seems in no way to restrict what the coin *could* do in the actual scenario. It still could have come up tails; it was a fairly weighted coin and so, as stipulated in Vihvelin's example, it had a 50-50 chance of coming up either heads or tails. Black's reliable prediction merely entails that it won't come up tails, not that it cannot.

Still, some will undoubtedly object. At first glance, there is certainly something funny about saying that the coin ever could come up tails when we have been given a scenario that is tailored to ensure that it never actually will. In one sense, the game is indeed rigged - I will always win. But on those individual instances when Black does not intervene in any way whatsoever, because he has predicted that the coin will come up heads all on its own by chance, it is hard to see how anyone could complain that the individual coin toss was unfair.

One might press the objection in this way, as Vihvelin describes: "If the coin were about to land tails, Black would have predicted this and intervened. And if Black had predicted this and intervened, the coin would be forced to land heads. So if the coin were about to land tails, it would be forced to land heads."[111] What this argument appears to show is that since Black counterfactually would intervene in any instance in which an uninhibited coin toss would have come up tails, it therefore follows that it cannot, for any toss, do otherwise than come up heads.

---

[111] Kadri Vihvelin, "Freedom, Foreknowledge, and the Principle of Alternate Possibilities", 20.

This argument initially seems convincing, but it runs afoul of a classic error - it is a hypothetical syllogism, which is generally accepted to be an invalid form of argumentation. This is a familiar point, which Vihvelin illustrates with a neat example: "If I jumped off this bridge, I would have arranged to be wearing a parachute. If I were wearing a parachute, I would not be killed. So if I jumped off this bridge, I would not be killed."[112] Clearly the first two counterfactuals are true, but the conclusion obviously does not follow. Another compelling example is provided by Stalnaker, and discussed by David Lewis: "If J. Edgar Hoover had been born a Russian, then he would have been a communist. If he had been a Communist, he would have been a traitor. Therefore: If he had been born a Russian, he would have been a traitor."[113] Again, the first two conditionals seem true (or at least very likely), and yet it is clear that even if they are true, the conclusion does not follow. The counterfactual syllogism used to argue that "if the coin were about to land heads, it would be forced to land tails" follows the same structure, and thus, Vihvelin concludes, that inference is invalid as well.

One way to reply to Vihvelin would be to argue that there is something distinct about the hypothetical syllogisms above that makes them invalid, whereas the kind of hypothetical syllogism used in a Frankfurt scenario is valid.[114] Fischer pursues a strategy of this sort, arguing that hypothetical syllogisms are invalid just in those cases in which the premises engage in "world hopping" – when they refer to counterfactuals that are true

---

[112] Ibid.

[113] David K. Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973), 32-33.

[114] The hypothetical syllogism in a Frankfurt scenario like the one I have been using might be something like this: If Riley were about to refrain from walking away, she would have blushed. If Riley had blushed, then Jesse would have intervened and forced her to walk away. Thus, if Riley were about to refrain from walking away, Jesse would have intervened and forced her to walk away.

in different possible worlds. For instance, in the Hoover example, the first premise refers to a possible world in which Hoover is a Russian, whereas the second premise apparently refers to a world in which Hoover is an American; thus there is no guarantee that there is any world in which the concluding counterfactual is true. Fischer argues that there is no problematic world hopping in Frankfurt scenarios, so there is no reason to think the hypothetical reasoning is invalid.[115] And Fischer argues further that, unlike the coin-flipping case, in standard Frankfurt scenarios the intervener is *on the scene*, ready to intervene. This, Fischer claims, makes the relevant counterfactuals in the Frankfurt-style argument true, and thus the hypothetical syllogism is sound.[116] And furthermore, Fischer adds:

> Additionally, I am not at all convinced that one needs to regiment the analysis as suggested by Vihvelin. Here are the facts in a Frankfurt case. There is a triggering event that occurs to indicate that the agent is about to refrain. The "counterfactual intervener" watches for that and even has the power and intention to intervene upon noticing the triggering event. Further, the counterfactual intervener is a completely reliable triggering-event detector and is completely reliable in carrying out his intentions. Given these facts, it just seems intuitively obvious that if the relevant individual (Jones) were about to refrain, he would be rendered unable to refrain. And thus it seems intuitively obvious that Jones is unable to do otherwise, given the facts of the case; no argument employing hypothetical syllogism or transitivity appears to be required.[117]

In short, in Fischer's view we don't even have to rely on any hypothetical reasoning. It is just intuitively obvious that the targeted agent in a standard Frankfurt scenario cannot do otherwise than he actually does.

In a later paper, Vihvelin replies to Fischer and argues that his responses are inadequate to rescue Frankfurt scenarios. She rejects the "world hopping" diagnosis of

---

[115] John Martin Fischer, *Deep Control* (Oxford: Oxford University Press, 2012), 62-62.
[116] Ibid., 60.
[117] Ibid., 65.

invalid hypothetical syllogisms. And she argues that putting the intervening agent "on the scene" makes no difference to the abilities of a targeted agent like Jones, and constructs some new and revised versions of her examples in an attempt to demonstrate this point. Here is one example:

> Jones is depressed and suicidal and Black is following him around to make sure he doesn't hurt himself. Right now Jones is perched on the ledge of a high building. If he falls, he will plunge instantly to his death. Black has a safety net, but it takes time to set it up; if Black waits until Jones jumps it will be too late to save Jones. Luckily for Jones, Black is able to reliably predict Jones's actions shortly before they occur, and this gives him enough time to get the safety net in place. Today Black correctly predicts that Jones will not jump, so he doesn't bring out the safety net. Query: what would have happened if Jones had jumped? Answer: he would have plunged instantly to his death (since the safety net was not there).[118]

Here Black is in some sense "on the scene", and yet it seems clear that at the moment in question (when the net is not set up), Jones retains the ability to leap to his death. And further, Vihvelin questions Fischer's appeal to intuition, arguing (rather fairly, it seems) that we should move beyond mere appeals to intuition to settle difficult modal questions like these.

This debate could of course be carried on further. We could question whether Black is really "on the scene" in the right way in Vihvelin's revised example (perhaps the gap in time it takes the set up the net is playing an important role here) and try to say more to defend Frankfurt scenarios that rely on counterfactual intervention. Perhaps such a strategy would be successful, but I think we may wind up in something of a dialectical stalemate. Instead, I prefer a different strategy of response to Vihvelin's argument – one that I think is more decisive, and does not rely on counterfactual intervention at all.

---

[118] Kadri Vihvelin, "Foreknowledge, Frankfurt, and Ability to Do Otherwise: A Reply to Fischer," *Canadian Journal of Philosophy* 38, no. 3 (2008): 360.

X. Blocking Vihvelin's Objection

The question to consider at this point is whether Vihvelin is right to claim that "conditional intervention" and "counterfactual intervention" actually exhaust all of the possibilities with regards to types of Frankfurt scenarios. The type of Frankfurt scenario I was defending before – blockage cases – do not seem to fit neatly in either category, at least not if they are properly described. Vihvelin in fact does briefly allude to blockage cases (of the sort described by Hunt and Mele/Robb) in her original 2000 paper. She suggests that such cases ought to be categorized as "conditional interveners". As she says, "imagine, as a recent variation on a Frankfurt story has it, that Jones has placed locks on the neural pathways in Jones's brain in such a way that, while Jones is never forced to choose as he does, had his deliberations taken any other course they - or rather their neurological realizations - would have found the alternative routes closed."[119]

In this way of describing blockage cases, Jones might attempt to deliberate otherwise than he actually does and find that something - the neural blockage - stands in his way. This way of describing the case certainly makes it seem to fit within the definition of a conditional intervener. But it's not clear that this is an apt way of describing the case. The point of the blockage cases is that it is not possible for Jones to even try to do otherwise, because all of the neural pathways that would be involved in his doing so would be blocked. Because of the blockages, Jones's deliberations cannot take any other course at all. The key characteristic of "conditional" cases is that "the conditional devices can only kick in if Jones makes a first wrong move."[120] But in a case where all of the neural pathways are blocked in advance, there is no first move that

---

[119] Vihvelin, "Freedom, Foreknowledge, and the Principle of Alternate Possibilities", 12.
[120] Ibid.

makes the "device" kick in. The device in this instance - the neural blockage - is already in place before Jones does anything at all.

Perhaps, then, blockage cases are better characterized as instances of counterfactual intervention. This seems closer, but still not quite right. Blockage cases are like counterfactual intervention in that the intervener acts prior to any action or choice that is contrary to the intervener's plans. But in the counterfactual cases that Vihvelin describes (and as they are usually described in the literature), the intervener waits for some earlier *sign* or *trigger* that indicates that the contrary choice is going to occur. In blockage cases, by contrast, the intervener waits for no such sign. In fact, in the scenarios described by Hunt, the intervener works by putting up the blockades in those instances in which the target agent is going to do as the intervener wants him to do, the opposite of what occurs in standard counterfactual scenarios.

So blockage cases do not seem to be easily characterized as either cases of conditional or counterfactual intervention, at least not as Vihvelin describes them. The question now - the most important question - is whether blockage cases succeed in depriving an agent like Jones of the ability to do otherwise than he actually does. I think it is very obvious that they do. If all of the alternate neural pathways that are necessary for *any* sort of deviation on the part of Jones have been blocked or removed in advance, then certainly Jones lacks the ability to deviate.

One might point back to the objection that neural blockage of the sort described by Hunt, insofar as it actually does guarantee a certain outcome, amounts to causal determinism, and thus cannot be used in constructing a legitimate, non-question begging Frankfurt scenario. Robert Kane has pressed this sort of objection, as mentioned before.

In my view this objection is misguided, as I argued in the earlier section. But at any rate, this is not an objection Vihvelin can help herself to. In her newer paper, she made things 'simpler' by telling a story in which determinism is true, so that it is uncontroversial that Black could make precise predictions about Jones' behavior (while also working from the assumption that Jones could sometimes do otherwise than he actually does). She did this to show that even in a case in which it is uncontroversial that Black could make perfect predictions, he is completely unable to alter any of the relevant modal facts about Jones. But blockage cases, as described here (and first developed by Hunt) clearly *do* allow an intervener like Black alter the modal facts about Jones - and they do so without altering any facts about his causal history or moral responsibility.

And if we start with the version of the story presented in Vihvelin's original paper - where Jones is an indeterministic agent, and Black has mysterious powers of perfect prediction - it seems clear that Black can cut of Jones' access to alternate possibilities via neural blockage without either causally determining his behavior or diminishing his responsibility. Vihvelin (or a defender of her view) could always try to press the point by arguing that perfect prediction is impossible, that there is something incoherent or impossible about such a Frankfurt story. But then this would simply mire Vihvelin in the traditional debate over Frankfurt stories, when what it seemed she wanted to do was demonstrate that Frankfurt stories failed *even if* you grant their supporters everything that they want. At the very least, one important lesson of blockage cases is this: anyone (like Vihvelin) who wishes to argue that moral responsibility is compatible with causal determinism is forced to admit that moral responsibility does not require access to alternate possibilities. This must be admitted, because a compatibilist can have no

principled objection to blockage cases, and yet it is very clear that they rule out any alternate possibilities for Jones, without in any way diminishing his moral responsibility.

XI. Conclusion

In this chapter I have argued that access to genuine alternate possibilities is not a requirement for moral responsibility. I did this by defending Frankfurt's strategy of arguing against PAP, arguing that the "blockage" strategy of the sort developed by David Hunt is the most promising route. I defended the blockage strategy by arguing that they presuppose nothing controversial other than the existence of future facts about what one will choose to do. I argued for various reasons that *this* is not a legitimate grounds for defending incompatibilism for various reasons, including an extended defense of the claim that nothing about the nature practical reason presupposes access to alternate possibilities. And I showed that blockage cases can be used to provide a strong response to Vihvelin's objection to Frankfurt's strategy, insofar as they clearly cut off access to alternate possibilities (and don't rely on any controversial hypothetical syllogisms). Ultimately, it is my view that if there is any legitimate grounds for incompatibilism, it will have to come from source worries rather than leeway worries. That will be the focus of the next chapter.

Chapter 4 – Source Incompatibilism

In the previous chapter, we explored one of the major strategies for defending the incompatibilist viewpoint, a strategy that has been called leeway incompatibilism. As we discussed, on this view the primary or fundamental explanation for why genuine freedom and moral responsibility is incompatible with causal determinism is that freedom and responsibility require access to genuine alternate possibilities, which causal determinism rules out. Some philosophers have even contended that leeway considerations are the only feasible threat to our freedom and responsibility. For example, Fischer writes: "there is simply no good reason to suppose that causal determinism in itself (and apart from considerations pertaining to alternative possibilities) vitiates our moral responsibility."[121] However I think that this claim is mistaken. While it's true that leeway considerations occupy a deep space in our pre-theoretic intuitions, I think that an even more profound, compelling, and fundamental threat to the compatibilist viewpoint can be derived from "source" or "causal history" arguments in favor of incompatibilism. Responding to that sort of incompatibilist viewpoint will be the focus of this chapter.

I. Mere Links

Being in control of ourselves, being the true source of our decisions and actions, matters a great deal to us. We care that we are able to determine our own fates (or at least the bit of our fates that depends on our choices), it matters to us that what we do is truly *up to us*. This fundamental intuition suggests a view of freedom and responsibility connected to self-control without undue influence from outside sources. One classic expression of this view is found in the closing argument of Clarence Darrow's famous

[121] John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control*, 159.

defense of Nathan Leopold and Richard Loeb, two wealthy young law students facing the death penalty for the carefully planned and brutal murder of a boy named Robert Franks. Darrow did not deny that Leopold and Loeb had committed the crime, convincing them to plead guilty. Instead, he sought to rescue his clients from the death penalty. Part of his strategy was to argue that in a moral sense, his clients were not truly blameworthy. In Darrow's words, "What has this boy to do with it? He was not his own father; he was not his own mother; he was not his own grandparents. All of this was handed to him. He did not surround himself with governesses and wealth. He did not make himself. And yet he is to be compelled to pay."[122]

The reasoning that Darrow appeals to here resonates, even if most of us would want to reject his ultimate conclusion. His clients were the products of the environments and their ancestry. Their upbringing, their genes, what they were exposed to in their educations (in another part of the speech, Darrow goes on at length about the undue influence that the writings of Nietzsche – who Darrow characterizes as a brilliant maniac - had on his clients, especially Leopold), all of these things determined the monstrous characters that Leopold and Loeb developed – and none of these things were within Leopold or Loeb's control. Once we look at things this way, we might start to (and certainly, Darrow is trying to convince us to) see Leopold and Loeb as little more than mere links in a long causal chain, and therefore not truly responsible for their horrible crimes.

---

[122] Clarence Darrow, *Attorney for the Damned* (Chicago: University of Chicago Press, 2012), 65.

This basic intuition has been developed and advanced by a number of philosophers to defend incompatibilist viewpoints. For example, Robert Kane claims that if an "action did have such a sufficient reason for which the agent was not responsible, then the action, or the agent's will to perform it, would have its source in something that the agent played no role in producing . . . ultimately responsible agents must not only be the sources of their actions, but also of the *will* to perform the actions."[123] Derk Pereboom expresses the source incompatibilist intuition in the following principle: "If an agent is morally responsible for her deciding to perform an action, then the production of this decision must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she has no control."[124] Elizabeth Anscombe expresses the view this way: "My actions are mostly physical movements; if these physical movements are physically predetermined by processes which I do not control, then my freedom is perfectly illusory. The truth of physical indeterminism is thus indispensable if we are to make anything of the claim to freedom."[125] Or as Laura Ekstrom puts it, "Since it is not up to me what happened in the distant past, and it is not up to me what the laws of nature are, if my current actions are the consequences of the past and laws, then my current actions are likewise not up to me."[126] This view can be expressed more formally in something like the following argument:

---

[123] Kane, *The Significance of Free Will*, 73.
[124] Pereboom, *Living without Free Will*, 47.
[125] G. E. M. Anscombe, *Metaphysics and the Philosophy of Mind.* (Minneapolis: University of Minnesota Press, 1981), 146.
[126] Laura Waddell Ekstrom, *Agency and Responsibility: Essays on the Metaphysics of Freedom* (Boulder, CO: Westview Press, 2001), 6. The passage quoted here is closely related to Peter van Inwagen's "Consequence Argument". However van Inwagen's

1. We are free, in the sense required for moral responsibility, only if we are the ultimate sources of our choices and actions.
2. If causal determinism is true, then all of our choices and actions are ultimately caused by things that are outside of our control (e.g. the distant past and the laws of physics).
3. If our choices and actions are ultimately caused by things that are outside of our control, then we are not the ultimate sources of our choices and actions.
4. So if causal determinism is true, then we are not the ultimate sources of our choices and actions.
5. Therefore, if causal determinism is true, then we are not free in the sense required for moral responsibility.[127]

II. Compatibilist Sourcehood

Compatibilists, of course, must reject this argument. One way is to reject the first premise of the argument – to argue that we can choose and act freely (in the sense required for moral responsibility), even if we are not the ultimate sources our actions. Compatibilists have developed accounts of self-control and self-determination meant to ground freedom and moral responsibility without requiring the falsity of causal determinism. For instance, hierarchical compatibilists (see Frankfurt[128] for a prime example) develop models of self-determination that claim that we are free so long as our desires and preferences have the right sort of structure. On Frankfurt's view, we can be understood to be the true "source" of our actions in the sense required for freedom just so long as our second order volitions "mesh" with our effective first order desires.

---

argument is formulated in terms of the ability to do otherwise than one actually does, arguing that since - given determinism - the ability to do otherwise than what one actually does would imply the ability to change things about the past or about the laws of physics, we don't have the ability to do otherwise than we actually do. Since I don't think that freedom in the sense of the ability to do otherwise than we actually do is necessary for moral responsibility (as argued in the last chapter), I am not dealing specifically with van Inwagen's formulation here. For more, see Peter Van Inwagen, "The Incompatibility of Free Will and Determinism," *Philosophical Studies* 27, no. 3 (1975), 185-99.

[127] This formulation of the argument is borrowed (with slight modifications) from Kadri Vihvelin, "Arguments for Incompatibilism" (Stanford Encyclopedia of Philosophy), March 1, 2011, http://plato.stanford.edu/entries/incompatibilism-arguments/.

[128] Frankfurt, *Freedom of the Will and the Concept of a Person.*

Other philosophers place more emphasis on our ability to react and respond to the rights sorts of reasons (see for example Dennett[129], Arpaly[130], Susan Wolf[131], and Fischer and Ravizza[132]). Fischer puts it in terms of what he calls "guidance control".[133] Fischer believes that an agent acts freely in the sense required for moral responsibility when her actions are produced by mechanisms (like the ordinary operation of practical reasoning) that would lead her to "choose and act differently in a range of scenarios in which she is presented with good reasons to do so."[134] As long as an agent is guided by appropriately reasons responsive mechanisms (and as long as she takes ownership of those mechanisms), we may understand her as having done things "her way" (Fischer alludes to Frank Sinatra's iconic song to illustrate the point).[135]

In addition to developing positive compatibilist versions of the sourcehood requirement on freedom and responsibility, many philosophers have expressed criticisms of the "ultimate" or "total control" versions of sourcehood that have driven incompatibilist thinkers. For instance, Nietzsche says:

> The *causa sui* is the best self-contradiction that has been conceived so far, it is a sort of rape and perversion of logic; but the extravagant pride of man has managed to entangle itself profoundly and frightfully with just this nonsense. The desire for "freedom of the will" in the superlative metaphysical sense, which still holds sway, unfortunately, in the minds of the half-educated; the desire to bear the entire and ultimate responsibility for ones actions oneself, and to absolve God, the world, ancestors, chance, and society involves nothing less than to be precisely this *causa sui* and, with more than Munchhausen's audacity, to pull oneself up

---

[129] Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting.*
[130] Arpaly, *Merit, Meaning, and Human Bondage.*
[131] Susan Wolf, "Asymmetrical Freedom," *The Journal of Philosophy* 77 (1980): 157-66.
[132] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.
[133] Fischer contrasts this with what he calls "regulative control", which involves genuine metaphysical access to alternative possibilities.
[134] Fischer, *My Way: Essays on Moral Responsibility*, 18.
[135] Ibid.

into existence by the hair, out of the swamps of nothingness.[136]

And Fischer expresses the compatibilist disregard for ultimate sourcehood or control nicely as follows:

> So total control is a chimaera. It is manifestly ludicrous to aspire to it or to regret its absence. The locus of control is not wholly within us. We do not exist in a protective bubble of control. Rather, we are thoroughly and pervasively subject to luck: actual causal factors entirely out of our control are such that, if they were not to occur, things at least might be very different. Quite apart from any special assumption about causal determinism, we can see that from a broader perspective, it is entirely a matter of luck or arbitrary that I behave as I do (or even that I developed into an agent at all — or have maintained that status). Although it is perfectly reasonable to wish to be the source of one's choices and behavior, it is not reasonable to interpret the relevant notion of sourcehood in terms of total control and internality.[137]

And even some libertarian incompatibilists lament the need to try to make sense of ultimate causation involving indeterminism. As Kane says, "Libertarians and incompatibilists do not want indeterminism for its own sake. If the truth be told, indeterminism is something of a nuisance for them. It gets in the way and creates all sorts of trouble. What they want is ultimate responsibility, and *ultimate responsibility requires indeterminism*. It has been said that all valuable things come with a price. In this case ultimate responsibility is the valuable thing, and indeterminism is the price. And indeterminism is a *high* price. For it threatens to subvert the entire incompatibilist project."[138] In this passage, Kane points to the fact that even if free will skeptics like

---

[136] Friedrich Nietzsche, *Beyond Good and Evil*, trans. Walter Kaufmann (New York: Random House, 1966), 28.
[137] Fischer in John Martin Fischer et al., *Four Views on Free Will* (Malden, MA: Blackwell Pub., 2007), 68.
[138] Robert Kane, "Two Kinds of Incompatibilism," *Philosophy and Phenomenological Research*, 2nd ser., 50 (1989): 227.

Nietzsche (see also Galen Strawson)[139] are right that it is impossible to make coherent sense of ultimate sourcehood or true self-causation, it could nonetheless be the case that genuine moral responsibility requires it.

At this point, it has seemed to some that we are left at a dialectical stalemate. Some people find the need for ultimate sourcehood compelling – so compelling that they are willing to pay the price of indeterminism, so compelling that many are willing to abandon our common sense ideas of freedom and responsibility altogether. But many others dismiss the idea of ultimate sourcehood as confused or absurd, and find it misguided to hope for it or try to make sense of it.

Given this sort of impasse, one might make the following suggestion – if source incompatibilists have nothing more to offer than an appeal to intuition that many of us do not share, and if this intuition is a threat to so much else that we intuitively hold dear, then on reflection it is an intuition that should either be discarded or revised in favor of something more tenable.[140] As Pereboom concedes, "one should not expect compatibilists to be moved much by this incompatibilist intuition alone to abandon their position."[141] The question now is whether incompatibilists can provide us with any strong reason for taking the ultimate sourcehood requirement for freedom and responsibility seriously. In the following sections I will explore one of the most compelling kinds of arguments in the incompatibilist's arsenal, and consider some possible compatibilist lines of reply.

---

[139] Galen Strawson, "The Impossibility of Moral Responsibility," *Philosophical Studies* 75, no. 1-2 (1994): 5-24.
[140] I will defend this claim in more detail in the last chapter.
[141] Pereboom, *Living Without Free Will*, 89.

III. The Problem of Manipulation

Thought experiments involving coercive manipulation have been used by a number of philosophers in recent years to undermine compatibilist accounts of freedom and responsibility, and to motivate the intuition that genuine moral responsibility does indeed require something like ultimate sourcehood. They are a potent piece of rhetoric in any case against compatibilism. One of the classic formulations of a manipulator argument was given by Michael Slote,[142] to reveal the inadequacy of hierarchical compatibilist accounts like Frankfurt's.[143] Slote's example involves a hypnotist who tinkers with an unwitting subject's second order volitions to produce a "free" (on Frankfurt's model) action. I will now describe an example that is similar in structure to Slote's.

Imagine a woman named Cameron who suddenly finds that she is struggling with inexplicable homicidal urges. She has been a calm and peaceful person for her entire life, but now Cameron frequently finds that she has a very strong desire to strike out and harm or kill people near her. She has never harmed anyone physically before, and reflectively – at the level of her second order desires – she prefers to keep things that way. Cameron is worried that the urges she is struggling with might one day become too strong to resist, so she seeks out the assistance of a hypnotherapist. Unfortunately for Cameron, the hypnotherapist she finds, Sarah, has her own agenda. Sarah – a devoted student of Frankfurtian thinking on free will – would like to bring Cameron's first and second order

---

[142] Michael Slote, "Understanding Free Will," *The Journal of Philosophy* 77 (1980): 149-50.

[143] For an earlier example of a manipulation argument targeting the simpler classic compatibilist account of freedom as uninhibited volition, see Richard Taylor, *Metaphysics* (Englewood Cliffs, NJ: Prentice-Hall, 1963), 45.

desires in line, so that Cameron might act freely. Sarah finds that it is much easier to alter Cameron's higher order aversion to murder than it is to eliminate her first order urges to kill. And anyway, Sarah finds the idea of making it so that someone else freely commits murder thrilling. So after an intense hypnotherapy session, Cameron now finds that she is completely rid of her reflective preference not to commit murder. Instead, Cameron now reflectively relishes the thought of acting on one of her first order homicidal urges. Before long Cameron finds herself in a situation where the urge strikes her, and she murders an innocent stranger.

Did Cameron act freely when she committed murder? And is she blameworthy for the murder? The overwhelming intuition is that Cameron was *not* acting freely, and that she is not morally responsible for the murder – or at the very least, that her responsibility is severely diminished. After all, before Cameron visited the hypnotherapist, she didn't *really* want to commit the murder. The homicidal urges were she was struggling against, not something she identified with or embraced. It was only after Sarah manipulated her that she identified with her homicidal urges. It seems that if anyone deserves blame for the murder, it is Sarah, not Cameron.

It is obvious how this kind of example threatens to undermine Frankfurt's account of acting freely. Cameron fully meets the conditions for freedom according to Frankfurt's model – she is an agent whose reflective second order volitions mesh perfectly with her effective first order desires. Someone who wishes to defend a hierarchical model like Frankfurt's might bite the bullet and say that since Cameron is now acting on the desires that she reflectively prefers to act on that she is free. But prima facie this seems much too large a bullet to bite; saying that an agent is free when she has been coercively

manipulated in the way that Cameron was (especially if we are trying to build an account of freedom to ground moral responsibility) seems too implausible. What's more, some incompatibilists argue that manipulation cases like this one can be developed to undermine *any* compatibilist account, showing that no compatibilist account is sufficient to ground claims of moral responsibility. I will now turn to one detailed and clever incompatibilist line of argumentation that attempts to demonstrate this.

IV. Pereboom's Four-Case Argument

Pereboom develops a version of the manipulation strategy that he calls the "four-case argument". Pereboom starts as above, describing an agent who meets all of the conditions of several of the most plausible compatibilist accounts of freedom and responsibility. For illustration, imagine a version of Cameron as described above – a rationally egoistic (but not purely egoistic – she sometimes acts on moral reasons, when they do not conflict too much with her own interests) who sometimes acts on her impulse to murder. Cameron's effective first order desire to murder conforms to her second-order volitions,[144] she is not constrained by any irresistible desire and her actions are not out of character for her,[145] her process of deliberation is moderately responsive to reasons in a way that displays a sane, stable pattern,[146] she retains a general capacity to grasp moral reasons and regulate her behavior by them[147] (though she sometimes chooses not to, such as when she decides to act on her impulse to kill), and we can add that she has a general capacity to gradually adjust and improve her character over time. According to most

---

[144] Frankfurt, "Freedom of the Will and the Concept of a Person", 1968.
[145] A. J. Ayer, *Philosophical Essays.* (London: Macmillan, 1954), 3-20.
[146] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, 1998.
[147] R. Jay. Wallace, *Responsibility and the Moral Sentiments*, 1994.

plausible compatibilist accounts of moral responsibility, Cameron is morally blameworthy for her act of murder, even if causal determinism his true. To subvert this claim, Pereboom asks us to consider four different variations on how Cameron's behavior might be determined by outside forces. I will summarize four similar cases using Cameron below.

In **Case 1**, Cameron is created by neuroscientists who can directly manipulate her brain through the use of radio-like technology. The scientists produce Cameron's mental states from moment to moment, pushing buttons to manipulate her reasoning process so that she is rationally egoistic in a way that leads her to sometimes act on her murderous impulses. The reasoning process that they give to her is responsive to reasons (including sometimes moral reasons), she has second order desires that mesh with her first order desires, and so on.

In **Case 2**, Cameron was created by neuroscientists. However these neuroscientists cannot control her directly from moment to moment; instead, they programmed her in advance to have the character that she has, a character that will lead to her occasionally acting (in a reasons-responsive, rationally egoistic way) on her reflectively endorsed desire to kill.

Case 3 does not involve any neuroscientists. Instead, we are to imagine that Cameron's rationally egoistic, moderately reasons-responsive, reflectively endorsed character is produced by rigorous and intensive home training and conditioning. This intensive conditioning took place at an age when Cameron was too young for her to have had the ability to prevent or alter or resist the conditioning in any way.

Finally, **Case 4** involves nothing more than ordinary physical causal determinism. Cameron had an ordinary upbringing (without any sort of rigorous conditioning or abuse or anything of that sort) and has developed and grown in an ordinary social environment which, together with her genes and past (and in general all the past of the universe plus the laws of physics) was causally sufficient to determine that she would wind up with the rationally egoistic, reflectively endorsed, reasons responsive, murderous character that she now possesses.

The aim of Pereboom's argument involving these four cases is to show that causal determinism is just as threatening to freedom and responsibility as coercive manipulation. Pereboom approvingly quotes Spinoza as follows, suggesting that this thought should "shape our reaction"[148] to manipulation examples: "Men think themselves free, because they are conscious of their volitions and appetite, and do not think, even in their dreams, of the causes by which they are disposed to wanting and willing, because they are ignorant."[149] By moving through the four cases, Pereboom argues, this truth should become obvious to us.

In the first case, the intuition that Cameron does not act freely and is not morally responsible is quite strong. Pereboom says that we have this intuition because the victim "is determined by the neuroscientists' activities",[150] which are completely beyond her ability to control. Pereboom then claims that we have the same intuitive reaction to Case 2, and he argues that this is the appropriate reaction. And again, Pereboom claims that the

---

[148] Derk Pereboom, "A Hard-line Reply to the Multiple-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77, no. 1 (2008): 161.
[149] Benedictus De Spinoza, *The Collected Works of Spinoza*, trans. E. M. Curley (Princeton, NJ: Princeton University Press, 1985), 440.
[150] Pereboom, *Living Without Free Will*, 113.

best explanation for this fact is that the manipulated agent is determined by factors outside of her control. Pereboom suggests that it would be an ad hoc, unprincipled move to suggest that Cameron "is morally responsible because the length of time between the programming and the action is great enough."[151] In Pereboom's view, the mere fact that the neuroscientists did the programming long ago in Case 2, as opposed to moment-to-moment programming, as in Case 1, should make no moral difference.

Pereboom then claims that generalizing from Case 2, we should have the same reaction to Case 3. The causal inputs are less "weird" than they are in Case 2, but there is no morally relevant difference between the two cases. In both cases, Cameron's character is shaped long ago by factors she could not control, thus we should conclude that Cameron in Case 3 lacks freedom and responsibility for much the same reason that Cameron of Cases 1 and 2 does. And this brings us at last to Cameron in Case 4. In Case 4, there is nothing very unusual about Cameron's story. But insofar as her character was brought about by causally sufficient conditions that she could not control, we ought to conclude that she – just like Cameron in Cases 1 through 3, is not morally responsible.

Pereboom's version of the manipulation argument presents a considerable challenge for the compatibilist. The compatibilist must either explain some principled relevant difference between Cameron in Case 4 and Cameron in the other cases, or else the compatibilist must find some way to embrace the result that we can act freely and responsibly even when coercively manipulated. One obvious move that the compatibilist might make is to point to the fact that in Case 4, unlike the other cases, Cameron's

---

[151] Ibid, 114.

character "is not, in the last analysis, brought about by other agents."[152] However Pereboom claims that the principle implied in this response is not sufficient to explain Cameron's lack of responsibility in the other cases. I will return to this point in more detail shortly, but first I would like to consider some other possible responses to Pereboom's version of the manipulation argument.

V. Does Determinism Explain Our Intuitions?

Pereboom's manipulation argument hinges on the claim that the best explanation for an agent's lack of responsibility in examples like Cases 1 – 3 is that her "action results from a deterministic causal process that traces back to factors beyond his control."[153] In response, Al Mele has argued that this conclusion is not warranted. Mele points out that we can construct close analogues of the first three cases that do *not* involves causal determination, and which nonetheless seem just as threatening to moral responsibility.[154] For instance, consider an analogue of Case 1 in which the neuroscientists control Cameron from moment to moment by pushing buttons that manipulate her brain via radio signals. However each time they push a button, there is a tiny probability that the machine will incapacitate Cameron instead of causing her to act as she is directed. Whether the machine causes her to act as directed or incapacitates her is truly random. Does this addition of indeterminism into the story change anything of our intuitions about Cameron's moral responsibility? It seems clear to me that it does not. Mele suggests a similar revision of Case 2, in which the programming that the

---

[152] Ibid, 115.
[153] Ibid, 116.
[154] See Al Mele, "A Critique of Pereboom's 'four-case Argument' for Incompatibilism," *Analysis* 65, no. 285 (January 2005): 75-80.

neuroscientists do when Cameron is young has a slight random chance of later incapacitating her rather than leading her to commit murder. Again, the added indeterministic element does not seem to make any difference.

Mele concludes that "for all Pereboom has shown, it is the manipulation, not the deterministic causation, that does the intuition-driving work in his cases."[155] And on this point he seems to be right. However this doesn't seem to undermine Pereboom's broader argument. Pereboom's ultimate aim, after all, is to prove the truth of something like his ultimate origination principle – that a necessary condition for an agent's moral responsibility is that it not be ultimately produced by a source over which she lacks control. In Mele's indeterministic versions of the cases, the same principle still serves to explain why the manipulated agent is not responsible. Deterministic manipulation is one way to bring about an agent's character outside of her control, but it is not the only way. As Pereboom says: "The point of the four-case argument is that determination by factors beyond one's control is sufficient for non-responsibility, for the reason that it precludes the kind of control required for moral responsibility. This point is consistent with the claim that there are other conditions, potentially the theme of other manipulation cases, that are also sufficient for non-responsibility. So determinism's not being essential to Plum's lacking moral responsibility does not undermine the argument."[156]

Still, Mele's point is important for at least one reason – it raises the possibility that manipulation arguments are not a problem uniquely for compatibilism. Given that manipulation can undermine the free will of indeterministic agents as well, it may be

---

[155] Ibid, 80.
[156] Derk Pereboom, "Defending Hard Incompatibilism," *Midwest Studies in Philosophy* 29, no. 1 (2005): 237.

possible to construct manipulation examples that are problematic for libertarian accounts of freedom and responsibility. This is a point I will return to in the next chapter.

VI. The Hard-Line Reply

Michael McKenna recommends that compatibilists be more willing to embrace what he calls a "hard-line" response to manipulation arguments like Pereboom's four-case argument. To begin, he draws a distinction between hard-line and soft-line responses.[157] A hard-line response is when a compatibilist accepts that a manipulated agent has met all of the necessary conditions for freedom and responsibility, and embraces the (at least prima facie) counterintuitive conclusion she is just as free as the rest of us. It is exemplified in this passage from Frankfurt: "A manipulator may succeed, through his interventions, in providing a person not merely with particular feelings and thoughts but with a new character. That person is then morally responsible for the choices and the conduct to which having this character leads."[158] Most compatibilists have found this kind of response unpalatable or untenable, and so they tend to favor a soft-line response to manipulation arguments. A soft-line response is when a compatibilist argues that a given manipulation example has failed to capture the compatibilist account of freedom it is meant to attack – that there is some morally relevant difference between the manipulated agent in the example and ordinary free agents acting under ordinary causally deterministic conditions.

---

[157] Michael McKenna, "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77, no. 1 (2008): 143.
[158] Harry Frankfurt, "Reply to John Martin Fischer," in *Contours of Agency: Essays on Themes from Harry Frankfurt*, ed. Sarah Buss and Lee Overton (Cambridge, MA: MIT Press, 2002), 27.

McKenna contends that while in some instances the soft-line response might be appropriate (as he says, "there is no one size fits all reply"[159] to manipulation arguments), taking a soft-line stance has an inherent problem – "it leaves open an easy incompatibilist rebuttal via a slight revision to the example"[160] so that the manipulated agent now meets all of the alleged compatibilist requirements of freedom in question. In other words, "a soft-line reply to a well-crafted version of MA can only temporarily forestall the inevitable."[161] He therefore recommends that compatibilists go on the offensive, so to speak, and take the hard-line response whenever possible. When a candidate manipulation example doesn't quite live up to the compatibilist conditions in question (or if it is ambiguous whether it does), McKenna recommends being charitable - fixing or clarifying the example so that it clearly *does* meet the relevant compatibilist conditions of freedom, and than embracing the conclusion that the manipulated agent is free and responsible.

VII. McKenna's response to the Four-Case Argument

With this recommendation in mind, McKenna sets his sights on Pereboom's four-case argument. To mitigate the intuitive cost of a hard-line reply in this instance, McKenna attempts to turn Pereboom's strategy against him. McKenna starts with something like Pereboom's Case 4, and calls attention to the manipulated target's agential and moral properties. So in my version of the example, we would reflect on facts like Cameron's capacity to grasp and act on moral reasons, her reflective endorsement of her murderous desires, her general responsiveness to reason, the stability of her character,

[159] Ibid.
[160] Ibid.
[161] Ibid, 144.

etc. When we reflect on the richness of Cameron's agential and moral capacities even in a deterministic environment, McKenna contends that it is at least not *clearly* the case that Cameron lacks freedom and responsibility. He suggests that open-minded observers should admit that the compatibilist account of freedom and responsibility is at least a *plausible* contender. It would be question begging for an incompatibilist to assume otherwise.

From here, we can move to Case 3. Using Pereboom's generalization strategy, McKenna argues that since Case 3 is in no relevant respect different from Case 4, we should conclude that it is not obvious that an agent like Cameron lacks freedom and responsibility in Case 3 either. And we should reason the same in moving from Case 3 to Case 2, and from Case 2 to Case 1. Hence, McKenna claims, we should have the same reaction to Case 1 that we do to Case 4 – that it is not obvious that an agent like Cameron in Case 1 lacks freedom and responsibility – we should treat Case 1 with the same open-minded agnosticism that we do Case 4. It is important to note here that McKenna is NOT using Pereboom's strategy to demonstrate the truth of compatibilism. Rather, he is merely seeking to diminish the force of manipulation examples as an argument FOR incompatibilism. If McKenna is right, then manipulation examples fail to prove that ultimate sourcehood is a requirement for genuine freedom and moral responsibility.

An obvious question to raise here is whether McKenna rightly characterizes the appropriate "agnostic" reaction to Case 4. Pereboom challenges McKenna on this point. He argues that the appropriate response to determinism:

> …affirms that determinism provides a reason for giving up the responsibility assumption, but claims that so far the issue has not been settled. Its advocate would say about an ordinary case of an immoral action, in which it is specified that the action results from a causally deterministic process that traces back

> beyond the agent's control, that it is now in question whether the agent is morally responsible. Call this the *neutral inquiring* response. By this response it is initially epistemically rational not to believe that that the agent in an ordinary deterministic example is morally responsible, and not to believe that he is not morally responsible, but to be open to clarifying considerations that would make one or the other of these beliefs rational.[162]

Pereboom agrees with McKenna that we should regard the issue as unsettled and in need of clarifying considerations when we first look at the Four-Case argument, but also is suggesting that prima facie there is a presumption of favor of incompatibilism. Pereboom then goes on to argue that the four-case argument is just the sort of clarifying consideration needed to prove the incompatibilist's case. In Pereboom's view, the generalization strategy used in the four-case argument, by showing that there are no morally relevant differences between the other cases and Case 4, bolsters the initial presumption that determinism is a threat to freedom, tipping the dialectical scales in favor of incompatibilism. This seems especially clear when we reflect on something as intuitively freedom-undermining as Case 1.

A related sort of question for McKenna's argument concerns the disparity in the kinds of conclusions that one gets depending on which way the argument is ran. Running from Case 1 to Case 4, we get (according to Pereboom) support for the conclusion that determined agents like Cameron *are not* morally responsible. Running from Case 4 to Case 1 (as McKenna suggests), the compatibilist friendly conclusion is much softer – that Cameron's responsibility hasn't been ruled out yet. It might be objected then that McKenna is unfairly holding the incompatibilist to much higher standards here. McKenna replies to this consideration by arguing that this is appropriate because of the

---

[162] Derk Pereboom, "A Hard-line Reply to the Multiple-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77, no. 1 (2008): 162.

bizarre, unnatural sorts of contexts the incompatibilist is relying on to pump our intuitions by use of manipulation cases. As he says, "The cases combined with our intuitive reacts to them must overwhelmingly speak in her favor. If any reasonable doubt can by cast on what our intuitions would tell us in these very bizarre contexts, then there is good reason to be unsure about how telling our own intuitions can be."[163]

This point seems fair – if the incompatibilist expects to prove his position with an appeal to intuition, then the intuitions should be very compelling ones. This is especially true when what is being argued for is threatening to something as deeply intuitive as moral responsibility. Still, I think we can bolster the compatibilist case further by focusing our attention on some additional pro-compatibilist considerations when we start with Case 4. I will now turn to some such considerations that I think help refine and strengthen the kind of hard-line reply McKenna is advocating.

VIII. Manipulation, Practical Reason, and Power Preferences

As I argued in the first chapter, drawing from the work of Hilary Bok, engaging in practical reasoning with moral ends gives us reason to distinguish between those actions that flow from, reflect, and shape our characters and those that do not, and to hold ourselves morally accountable – to see ourselves as apt targets of moral criticism – for the former. I also mentioned that this insight helps point the way to part of a response to manipulation arguments. In particular, I think it can help us to bolster the hard-line reply in some versions of the manipulation argument, and help further explain why a soft-line

[163] Michael McKenna, "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument", (2008): 157-58.

reply is appropriate for others. I would like to explore these points in some more detail here, drawing some further insights from the recent work of Keith Lehrer.

As I mentioned in the first chapter, as agents who engage in practical deliberation about how to act or what to do, we don't *just* decide on particular courses of action. In choosing how to act, we endorse certain standards for action and we reject others; we deem some goals and values and reasons as more worthy than others. If hold moral values to be important, and if through our deliberations and choices we fail to uphold our moral values, then we will hold ourselves accountable for that failure in a way that has a distinct emotional component (see the discussion of guilt in the first chapter). But this kind of reaction only seems fitting if our choice of action and our choice of standards is truly our own. If we were to learn that someone else *imposed* our choices on us in a coercive or manipulative way, then it seems that such a reaction would no longer be apt.

Lehrer calls attention to and elucidates this feature of practical agency in his account of autonomy. As mentioned before, a number of accounts of freedom and responsibility highlight the centrality of our being guided by reasons – including Bok's, and most notably the account developed by Fischer and Ravizza. But this is not all there is to a full account of free practical agency. As Lehrer says, "Fischer and Ravizza have insisted on the importance of being open to the guidance of reasons. They should recognize that whether something is a reason that will guide my actions – that is, a reason for me – depends on my preference for being guided by such reasons."[164] The key point here is that an agent who is receptive and responsive to reasons could still be an agent

[164] Keith Lehrer, *Art, Self and Knowledge* (Oxford: Oxford University Press, 2012), 95.

who is *manipulated* in a way that subverts freedom and responsibility. As Lehrer goes on to explain:

> Being guided by reasons in a way that renders me free and not manipulated by a line of reasoning requires that I choose, or prefer to choose, to be guided in my choices and actions by such reasons. It must be up to me, not only that I do what I do, but that I am guided by the reasons I am. Manipulation by reasons will not make me free. Suppose I am an artist whose work is manipulated by the reasoning of a successful gallery owner imposing his commercially viable aesthetic using financial pressure. My work is not a free action of self-expression. Obviously, if the free action is to be an action of self-expression, it must be up to me what reasons guide my actions.[165]

What Lehrer describes here can be seen as a sort of "source" requirement for freedom. It's not enough to be moved by reasons – we have to be moved by them because we prefer to be moved by them; the choice to be moved by the sorts of reasons that move us has to be ours.

However Lehrer is also quite clear that his version of the source requirement for freedom is meant to be compatibilist friendly. Lehrer develops a hierarchical model of autonomy and self-ownership that is reminiscent of Frankfurt's, but distinct in important details. Like Frankfurt, Lehrer notes that a key element of autonomy is that we have preferences in favor of the system of desires and reasons that move us. Lehrer further introduces the notion of a power preference – a preference for the set of preferences that one has, including the power preference itself. This introduces a referential loop into the account of autonomy, avoiding a problem of regress that has been raised for Frankfurt's account.[166] It also allows room for one to act autonomously with conflicts between

---

[165] Ibid, 95-96.

[166] To put it simply, the worry for Frankfurt's account is that if acting on a desire autonomously or freely requires second level endorsement, then what of the freedom of that second level desire? Does it require endorsement at a third level? And so on? For

desires and reasons within one's preference structure, just so long as one has a preference for a preference structure containing such conflicts. However having a power preference in favor of one's preference structure doesn't guarantee autonomy – one could still be subversively or coercively manipulated into having the preference structure that one has, including the power preference. For this reason, Lehrer introduces a loop of explanation as a further condition on the power preference. As he says, "we must add the further condition that I have the power preference because I prefer to have it."[167] This condition is meant to be consistent with causal determinism – "of course, there may be a chain of causes, but the power preference must be the primacy explanation."[168]

This is only a very brief sketch of some of the features of Lehrer's rich and detailed account of freedom and autonomy. But I think it is enough to highlight the central lesson that I wish to take from Lehrer, which is that a key component of free practical agency – the kind that can ground moral responsibility – is that we be able to act on the basis of reasons and goals that we have because we prefer to have them, not because they are *imposed* on us by another agent, because someone *else* prefer that we have them. As Lehrer notes, this is consistent with there being a causal history behind my coming prefer the reasons and desires that I act upon. What is important is that nothing in the causal history be the *primary* explanation, that it not subvert or overshadow the explanation in terms of my preference for my preference structure.

The incompatibilist is likely to object that if there is a deterministic causal chain leading up to the formation of my preferences, including my power preference, then my

---

early discussion of this objection, see Gary Watson, "Free Agency," *The Journal of Philosophy* 72 (April 1975): 205-20.

[167] Lehrer, *Art, Self, and Knowledge*, 183.

[168] Ibid, 184.

power preference cannot be the primary explanation of itself and my other preferences. Instead, something else in the deterministic causal chain (or perhaps the causal chain itself) would have to be the primary explanation. But it is not immediately obvious why we should accept this. Once an agent has a mutually self-supporting and self-explaining system of preferences is in place and operating free of coercive influence, then it seems like facts about the causal history that led up to that system will (at least in typical circumstances) fall away in explanatory relevance. Lehrer's metaphorical use of a keystone in an arch is enlightening here. The keystone in an arch is the last stone put in place, the central stone that makes it possible for the arch to support itself and to bear weight. "My choice of a keystone as a metaphor is carefully chosen because of the natural way in which the keystone, while it supports the other stones in the arch, is at the same time supported by this stones."[169] This is analogous to the power preference for one's own preference structure; the power preference allows the structure to be self-supporting, and it itself is supported by the other preferences.

Once we fix in mind the self-referential and self-supporting nature of our preference structure, then the incompatibilist intuition that a determinist causal chain subverts or takes primacy in explaining an agent's preference structure is weakened. The deterministic causal history is analogous to the scaffolding used to support the arch before the keystone is put in place. The scaffolding is a part of the causal history of the arch, but once the arch is assembled and the scaffolding is removed, the scaffolding loses its explanatory relevance. The arch is now in a significant sense independent of the scaffolding; a full explanation of the roles the structures of the arch play in supporting

[169] Ibid, 92.

itself and supporting whatever is around it need not appeal to the scaffolding that was once there. Very similarly, once our preference structure is in place, supported by and supporting the power preference that explains and endorses the structure, which in turn grounds and explains the choices and actions of the agent they belong to, the various details of the causal physical laws and the long causal history that led to the existence of that agent largely fall away like the scaffolding in explanatory relevance. If you want to explain that agent's choices and actions, you do so by reference to her preference structure; the causal history is largely (at least in most cases) irrelevant.

However there do seem to be some exceptions, some instances in which the features of an *unusual* causal history can figure prominently into our explanations of why someone acts as they do, and perhaps even take primacy. As Lehrer puts it (in reference to the autonomous creation of a work of art), "sometimes someone else places the keystone arch of meaning of an artwork and not the artist."[170] From our standpoint as practical agents, if our preference structure – in particular our power preference – is put in place by another agent in the wrong sort of way, then it is in a very real sense no longer *ours*. It would no longer make sense for us to regret or to take pride in the actions that flow from our preference structure. The manipulation argument, in particular Pereboom's version, can be understood as a strategy for showing that the considerations that make aspects of the causal history relevant in manipulation cases also make the causal history relevant in ordinary instances of causal determinism. This is still a bit sketchy, but with this general framework in mind, I think we can now begin to develop

[170] Ibid, 78.

some constructive comments to supplement McKenna's account of when it is appropriate to take hard-line or soft-line stances in response to cases of manipulation.

IX. The Relevance of Other Agents

Incorporating the above insights from Lehrer's theory into my framework, I think we can now more plausibly assert that it makes an important difference whether an agent's causal history includes the intentions and goals and preferences of another agent. It also makes a difference just what role the intentions and goals and preferences of the other agent played, and how pervasive the influence is. I will treat each of these points in turn.

First, let us return to Pereboom's Case 1. Recall that in Case 1, the manipulated agent (in my case, Cameron) is controlled from moment to moment by neuroscientists who push buttons that send radio waves to manipulate her brain. And recall that McKenna suggests that we ought to generalize from Case 1 and take a hard-line reply, asserting that it is not obvious that Cameron is not free in this case. This seems like a tremendous bullet for the compatibilist to bite. And with the framework laid out above, I think that we can see that it is a completely unnecessary bullet for the compatibilist to bite. It is clear that Cameron's choices regarding her preference structure are *not* the primary explanation of that structure, including her power preferences – her preferences and choices are constantly being put in place by *other* agents, acting according with *their* own preferences, reasons, and desires. It is the preference structures of the neuroscientists in control that primarily explain Cameron's choices and actions, so there is no reason for a compatibilist to bite the bullet and accept the claim that Cameron is a free and

responsible agent. In short, contrary to what McKenna has suggested, a soft-line reply seems warranted for Case 1.

McKenna suggests that we can pursue his strategy of slightly revising Case 1 to create a version that meets any plausible compatibilist conditions on freedom and responsibility. But it is not clear to me that this can be done, at least not in a way that would be satisfying to the incompatibilist. Here is what McKenna suggests about Pereboom's version of Case 1 (involving a manipulated agent named Plum):

> This team of neuroscientists, let us call them Team Plum, has a host of restrictions as to what they can do and what they can control about Plum, restrictions driven by the demands of authentic agency. Plum, for instance, must have an internally coherent and properly causally integrated mental life. His memories about past considerations must be able to inform and causally influence his current deliberations. And he must be causally linked to the external world in the proper way. If a bus is careening along out of control ready to hop up on the sidewalk and crush him, he is able to respond to those facts and leap from danger, and so on. Team Plum could be working from an elaborate control center orchestrating the various causal inputs that are involved in Plum's interactions with his world. On this model, while Team Plum is able to steer Plum in certain directions (like to kill Ms. White), often times, Team Plum is functioning merely as a sort of extra causal link in a chain. Team Plum functions like a prosthetic, allowing Plum to deal with his world like any other agent. So, let us suppose that Team Plum does not operate by taking Plum, as Mele puts it, ''out of the control loop.'' Let us instead assume that Team Plum operates by providing a very weird causal prosthetic, a causal foundation for the constitution of Plum's control (i.e., a foundation different from the foundation provided by typical neural realizers found in normal agents).[171]

I must confess that I am not completely sure about how to interpret McKenna's suggestion here, which is why I post it in its entirety rather than summarizing it. One of the aims of McKenna's suggested revision is clear at least – it is meant to help Pereboom's Case 1 to avoid some of the soft-line objections that some philosophers have raised. As noted in the passage above, Mele asserts that the reason that Plum lacks

---

[171] Michael McKenna, "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77, no. 1 (2008): 149-150.

freedom is that Plum is completely "out of the control loop",[172] because he is "not even partly in control of 'the process of reasoning' that happens in his head. Rather, 'his every state from moment to moment' is directly produced…by the neuroscientists."[173] So at least part of what McKenna seems to be up to in his constructive adjustment/clarification of Pereboom's Case 1 is putting Plum at least *partially* back in the control loop – making it so that existing facts about Plum (such as facts about his mental life – presumably including his preferences and desires and beliefs) play a role the choices Plum makes and how he acts. This much makes sense. But the bit about making the neuroscientists function "like a prosthetic", or "merely as a sort of extra causal link in a chain" makes the example less clear. If all that the neuroscientists do is orchestrate some of the causal factors that lead Plum to act, then it starts to sound more like the neuroscientists are just *influencing* him rather than actually controlling him. If that's the case, then perhaps it's true that the compatibilist ought to admit that he is morally responsible – but now it is much less clear why this is a difficult bullet for the compatibilist to bite, or why the example stands as an objection to compatibilism. Intuitively, it is not so difficult to grant that an agent can be morally responsible for an action when he is just influenced, but not controlled, by the actions of someone else. It seems clear that if the example is to retain its intuitive force, it needs to be the case that the team of neuroscientists manipulating Plum really is "directly producing his every state from moment to moment",[174] as Pereboom's original example stated. And we can add to this the additional requirement that how the neuroscientists can direct Plum is constrained by facts Plum's character and

---

[172] Mele, "A Critique of Pereboom's Four-Case Manipulation Argument", 78.
[173] Ibid, 77.
[174] Pereboom, *Living Without Free Will*, 113.

history, to make sure that he truly counts as an agent (and not a *mere* automaton or puppet) and is not *completely* out of the control loop.

Now that we've gotten a little more precise about what the structure of Case 1 ought to be, we can better see why it is such a powerful piece of rhetoric in supporting the source incompatibilist viewpoint. As should be clear now, the case can be described so that the manipulated agent like Plum (or Cameron in my earlier versions of the examples) meets the conditions of many of the most plausible compatibilist accounts of freedom, as described earlier (his actions can be guided by practical deliberation receptive and responsive to moral reasons, his second order desires can mesh with his first order desires, etc.). And yet, intuitively, it seems overwhelmingly obvious that an agent like Plum who is manipulated from moment to moment by a team of neuroscientists is *not* free, and he is clearly not morally responsible for his actions. McKenna recommends taking the hard-line response for this sort of case, as we saw above, but in this case the hard-line response seems to involve biting much too large a bullet. What is the compatibilist to say?

In my view, the correct compatibilist response is a soft-line response. The compatibilist should deny that Plum is a free and responsible agent. As I suggested before, drawing on my account of what is involved in moral practical reasoning (incorporating the insights from Lehrer's account of autonomy, as described above), I think the explanation for why Plum in Case 1 is not free is because his mental states from moment to moment can be directly attributed to actions of *other agents* (the neuroscientists). Plum cannot take proper ownership of his preferences and reasoning and choices because those features of his character belong to *others*, in the sense that they are

fully up to others. In Lehrer's terminology, the preferences and choices of the neuroscientists are the *primary explanation* of why Plum has the preferences that he does and chooses and acts as he does; they are the ones that put his "keystones" in place.

Some philosophers (like Pereboom and Mele) have objected to the idea that the fact that the manipulation is being done by other agents could be the feature of cases like Case 1 that explains why the manipulated agent is not responsible. They argue that we can change Case 1, replacing the agents with something non-agential, and the case remains just as threatening to freedom and responsibility. For illustration, Mele describes an alternate version of Case 1. "Imagine a variant of case 1 (case 1b) in which a strange, spontaneously generated electromagnetic field he is passing through on a cruise ship directly produces Plum's 'every state from moment to moment'".[175]

The issues here are delicate and tricky, but ultimately I think that Pereboom and Mele are mistaken in their claim that the replacement of agents with blind forces makes no difference in this case. To see this, we need to follow McKenna's advice, and make sure the cases are being described carefully so that it is very clear that Plum satisfies the relevant compatibilist conditions for freedom. Once we do this, then I think that a soft-line reply *is* appropriate in the revised versions of Case 1.

Let us consider Mele's variant case, Case 1b, in more detail. As Mele describes the case, I agree with him that it is plausible to claim that Plum is not morally responsible. But this is because the way Case 1b is described, it appears that Plum fails to satisfy some basic compatibilist conditions for moral responsibility. The fact that in this case Plum is manipulated by an electromagnetic field that he passes through on a cruise

---

[175] Mele, "A Critique of Pereboom's Four-Case Manipulation Argument", 77.

ship suggests that the field causes Plum's character undergo a change from what it was before he was on the cruise ship. But many philosophers (including Mele) insist that a necessary condition for moral responsibility is that one's own history of deliberation and reasoning and choosing shape one's current character. On such accounts, a radical change to one's character caused by an outside force, completely removed from one's process of deliberation and choosing, would undermine responsibility. If Mele's example is going to meet all of the plausible compatibilist conditions of moral responsibility, then it need to be revised.

So instead of imagining that Plum is suddenly changed by an electromagnetic field he is passing through, let's imagine the electromagnetic field has always been with him. Throughout Plum's causal history, the electromagnetic field has always played a causal role in the shaping and development of character from moment to moment – a character which is stable, responsive to moral reasons, has the right hierarchical structure, etc. Call this case 1c. Is Plum in Case 1c morally responsible for his actions? I think now the answer is yes, or at least that it is *plausible* that it is yes – the compatibilist doesn't have to bite a big bullet to accept Plum's responsibility here. Plum 1c is certainly quite unusual – his character, mind, preference structure, etc. is not physically realized in just a brain, but also partially in an electromagnetic field. But as long as Plum 1c's practical reasoning operates in a normal way and he has the right sort of preference structure and history, then there is no obvious reason why we cannot say that he is morally responsible for what he does. The only substantial difference between Case 1 and Case 1c is that the latter doesn't not involve manipulation by any agents – and yet I think it is clear (or at least plausible) that this is a difference that makes a difference.

To be clear, I don't mean to suggest that the presence of manipulative agents in standard manipulation cases is always what explains why the manipulated agent is not morally responsible. There can be cases in which agents do play a manipulative role *without* undermining responsibility, and there can be cases of "manipulation" with *no* manipulative agents that *do* undermine responsibility. My claim is just that agential manipulation *can* be the factor that explains why an agent is not responsible, in those cases in which the manipulation plays a causally explanatory role that takes primacy over the agent's own keystone states. Recognition of this fact will help compatibilists avoid taking the hard-line reply with regard to kinds of manipulation cases that don't warrant such a reply, like Pereboom's Case 1.

X. Pervasiveness and Sorites

If what I have argued so far is correct, then a soft-line reply is available to compatibilists for cases like Pereboom's Case 1, due to the fact that other agents are causing overwhelming moment-to-moment manipulation of the agent in question. The question now is what should we say about something like Case 2? The response again is going to be tricky and delicate, perhaps even more so than with Case 1. It will depend on just precisely how the case is described; the devil is in the details. As Pereboom describes Case 2, it is meant to be very similar to Case 1. The only significant difference between the two cases is supposed to be that instead of doing the programming from moment to moment, as in Case 1, the neuroscientists do all of the programming long in advance, say when the manipulated agent is very young. Pereboom claims that this difference cannot make any moral difference; "Whether the programming takes place two seconds or thirty

years before the action seems irrelevant to the question of moral responsibility."[176] But is this right?

Many philosophers advocate a soft-line reply to cases like Case 2. For example, Bernard Berofsky challenges the idea that a manipulator could remotely manipulate an agent into committing an action at a time much later in his life without violating some plausible compatibilist conditions of freedom, like the capacity to adjust our values and preferences over time in response to various unpredictable environmental influences.[177] In other words, Berofsky suggests that the most feasible way for a team of neuroscientists like the ones described in Pereboom's Case 2 to guarantee that a manipulated agent like Plum does exactly what they want him to do is to design him so that the features of his character that will lead to the desired action (e.g. his murderous impulses, his rationally egoistic value system, etc.) are immune to revision, no matter how he might later deliberate or what circumstances he might find himself in. But if an agent like Plum is manipulated in this way, so that he now lacks any general capacity to grow and shape his character via his practical deliberations and choices, then he clearly fails to satisfy some of the basic conditions of freedom and responsibility that have been proffered by many compatibilists. In short, understood this way, a soft-line reply to Case 2 is relatively easy. Mele seems to have a similar reading of cases like Case 2 in mind when he says of them, "Plum played no role at all in shaping his procedure for weighing reasons (say, through trial and error over the years he has been in the business of deliberating). Unlike normal

---

[176] Pereboom, *Living Without Free Will*, 114.
[177] Bernard Berofsky, "Global Control and Freedom," *Philosophical Studies* 131, no. 2 (2006): 427-28.

agents, Plum had no control throughout his history as an agent over this important aspect of his deliberative style."[178]

As he did with Case 1, McKenna argues that we can help Pereboom out, strengthening Case 2 so that it meets all of the relevant compatibilist requirements. He says, "imagine that the egoistic values that Plum came to acquire were the upshot of years of studying various ethical texts and an eventual considered fondness for the writings of Hobbes. Furthermore, these values were tested against others over the course of many years and various experiences led him to give up his ''experiments'' with others and eventually come to the egoistic values that informed his decision to kill Ms. White. In Case 2, the manipulators manipulating from a temporal distance will have a tough time pulling this off, but this is how they have to hit their mark."[179] The way McKenna describes the revised case, it seems that the manipulative power of the neuroscientists is pretty limited – similar to the influence that the neuroscientists possess in McKenna's revised version of Case 1. He later refers to their success in manipulating Plum to have just the right character to kill Ms. White as "dumb luck".[180] They seem able to influence the way he develops his character (by insuring that he reads the right texts, has the right experiences, etc.), but as McKenna describes it, they don't seem to have complete control.

Described this way, a soft-line reply again seems plausible. And it is understandable why McKenna opts for this sort of revision. He wants to make it clear that Plum satisfies any plausible compatibilist requirements for freedom and responsibility,

---

[178] Mele, ""A Critique of Pereboom's 'four-case Argument' for Incompatibilism", 78.
[179] McKenna, "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument", 151.
[180] Ibid, 153.

including the capacity to shape and refine his own character over time through his processes of deliberation and choice. If the team of neuroscientists manipulating Plum is going to leave this general capacity in tact, then it might seem that the neuroscientists that 'program' him are limited to influencing him in the sorts of ways that McKenna describes.

An interesting point emerges if Case 2 is understood in the way that McKenna suggests. Recall that Pereboom argues that the temporal difference between Case 1 and Case 2 cannot possibly make a moral difference. As he puts it, "it would seem unprincipled to claim that here, by contrast with Case 1, Plum is morally responsible because the length of time between the programming and the action is great enough."[181] But now I think we can begin to see how there *is* a principled reason for saying that this difference makes a moral difference. The reason is that, on plausible readings of Case 1 and Case 2, the fact that in the latter case the neuroscientists are temporally removed from Plum's actions seems to limit the amount of control they can have over Plum's actions without violating some plausible compatibilist requirements for freedom, like the ability to shape his own character via the operation of practical deliberation.

Another interesting related point that emerges here concerns a possible objection to Pereboom's approach – that the four-case argument is nothing more than an illicit sorites strategy. Pereboom references this possible objection and quickly dismisses it as follows:

> Notice that this generalization strategy is not a sorites. Its force does not depend on producing a series of cases, each of which is similar to its predecessor, and then arguing that since the first has some general feature, one must draw the conclusion that the last does as well because each of the successive pairs of cases

[181] Pereboom, *Living Without Free Will*, 114.

> is different only in some small degree of that kind of general feature. A series of similar cases is indeed important to the argument. But its strength derives from the fact that between each successive pair of cases there is no divergence at all in factors that could plausibly make a difference for moral responsibility, and that we are therefore forced to conclude that all four cases exhibit the same kind and the same degree of an incompatibilist responsibility-under- mining feature.[182]

In short, Pereboom rejects the idea that his four-case argument is a sorites because the only factor that could make a difference to moral responsibility – that Plum's behavior is causally determined by factors outside of his control – does not vary in any of the cases.

But if what I am suggesting is right – that the pervasiveness of the manipulator's influence and her level of control can make a moral difference – then the four-case argument *does* start to look like a sorites. As we move from Case 1 to Case 2 - and we can imagine a range of intermediate cases, where the team of neuroscientists become gradually more distant and their control gradually less direct and pervasive – the fact that explains Plum's lack of moral responsibility is changed "in some small degree", just like a sorites example. This same point holds as we move from Case 2 to Case 3 (and again, we can imagine a series of intermediate cases) - the role that the manipulators is made less direct and less pervasive, they have less control. To put it another way, the primacy of the explanatory role played by their intervention is diminished. And as we move from Case 3 to Case 4 (again, imagine intermediate cases) the role of the manipulators is gradually eliminated entirely.

If we imagine a long chain running the spectrum of cases ranging from Case 1 to Case 4, there won't be a clear dividing line between those cases in which Plum is responsible and those in which he isn't – just like in a typical sorites example. In other words, there won't be a sharp point at which the compatibilist ought to switch from a

---

[182] Ibid, 116.

soft-line reply to a hard-line reply. Instead, there will be a fuzzy boundary (I think somewhere between Case 1 and Case 2, on the understandings of those cases that I outlined above) where there will be difficult borderline cases. The central point here, of course, is that reasoning from Case 1 to the conclusion that Plum lacks moral responsibility in Case 4, on the basis of the fact that there is no clear sharp point on the spectrum at which we can say that he becomes morally responsible, would be just the sort of illicit reasoning that marks typical sorites arguments.

XI. A Different Reading of Case 2

The preceding comments depend on reading Case 1 and Case 2 in the ways that are described above. I think the incompatibilist will want to understand Case 1 in the way I suggested – it is the only way to generate the strong intuition that Plum is not morally responsible. However perhaps the incompatibilist will not be satisfied with the reading of Case 2. The incompatibilist might want to develop a version of this case that satisfies all of the relevant compatibilist conditions while also allowing the control to be more pervasive, as pervasive as it is in Case 1. To accomplish this level of control from a distance, it seems we need a manipulator greater than a team of neuroscientists.

McKenna actually describes a different case that fits the bill. Mele develops a similar example to motivate an argument for incompatibilism – he refers to it as the "zygote argument".[183] In these variations of the case, the manipulative team of neuroscientists is replaced with a manipulative deity named Diana. In Mele's version of the argument, Diana creates a zygote with absolute precision, using her perfect understanding of the deterministic laws of nature and the prior state of the universe to

[183] Alfred R. Mele, "Manipulation, Compatibilism, and Moral Responsibility," *The Journal of Ethics* 12, no. 3-4 (2008): 278-83.

arrange the atoms of the zygote in just such a way so as to ensure that in 30 years it will grow to be a person who freely commits some action that Diana wants committed – say Plum's murder of Ms. White. Further, Diana ensures that Plum will be an ideally rational agent who is receptive and responsive to moral reasons, who is rationally egoistic, who has the capacity to alter and improve his character over time, who has a preference for his preference structure, etc. We can understand Diana as a sort of uber-version of the team of neuroscientists, with absolute control over every detail of her target's actions, in spite of the fact that she does her work from a great temporal distance.

What should a compatibilist say about this sort of case? McKenna and Mele are in agreement – to them it seems obvious that a compatibilist ought to take a hard-line stance; she ought to assert that Plum, manipulated by Diana, is as free and morally responsible as anyone who is subject to ordinary causal determination. In McKenna's view, "it seems arbitrary to make theological determination itself have a relevant difference here."[184] Similarly, Mele questions "how it can matter for the purposes of freedom and moral responsibility whether, in a deterministic universe, a zygote…was produced by a supremely intelligent agent with Diana's effective intentions or instead by blind forces."[185] While they agree on this point, they seem to disagree about the significance of it. McKenna apparently takes this to be an unproblematic result for the compatibilist, a relatively easy sort of case about which to take a hard-line reply.[186] Mele,

---

[184] McKenna, "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument", 153.
[185] Mele, "Manipulation, Compatibilism, and Moral Responsibility", 280.
[186] One thing that makes this evident is that in his own series of cases (McKenna uses 6, as opposed to Pereboom's original 4), McKenna's two 'theological determination' cases are the closest on the spectrum to ordinary causal determinism – see McKenna, "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument", 152-53.

on the other hand, treats this case as problematic for compatibilism; in his view, such a case is a "significant part of what prevents some of us from coming down off the fence and endorsing compatibilism".[187]

In my own view, it is not obvious that a compatibilist is committed to taking a hard-line reply about cases like this one. From the framework I have outlined so far, it strikes me as plausible to say that the explanatory role that it played by an omnipotent deity is more threatening than the explanatory role played by ordinary causal determinism. This is a point I will return to in a little more detail in the next chapter, when I discuss a new and novel version of the manipulation strategy for defending source incompatibilism, one that involves theological determinism.

XII. Conclusion

This chapter has been dedicated to examining and criticizing the incompatibilist claim that we must be the *ultimate* sources of our own characters and actions (in a way that rules out causal determinism) in order to act with true freedom and moral responsibility. I argued that while there is something plausible about the demand for *some* sense of sourcehood, there is no principled reason for saying that the sourcehood requirement is as demanding as the incompatibilist claims. The bulk of the chapter was devoted to what I take to be the most compelling line of argumentation for an ultimate sourcehood requirement – manipulation arguments, best exemplified by Derk Pereboom's four-case argument. In line with McKenna, I argued that some versions of manipulation examples can be embraced by compatibilists without suffering any devastating costs to their position. And borrowing from the work of Keith Lehrer, I

[187] Mele, "Manipulation, Compatibilism, and Moral Responsibility", 283.

argued that a compatibilist-friendly account of sourcehood could be developed that could be used to help the compatibilist resist biting the bullet in some of the more problematic cases. Ultimately, I conclude that standard versions of the manipulation argument for incompatibilism have failed to provide compelling reason to accept anything like an ultimate sourcehood requirement for freedom and responsibility – and in my view, this is enough for the compatibilist to claim victory.[188]

In the next chapter, I will examine and critique two different recent and novel attempts to argue against compatibilism. One relies on the notion of prepunishment (which is the idea of punishing people for crimes that they have not yet committed, but which they will commit in the future), and the other is a very clever variation of the manipulation argument involving theological determinism. The two arguments are distinct, but in an interesting way related, insofar as they both point to difficulties that compatibilists allegedly have in talking about our attitudes towards people on the basis of causally determined actions that have not yet occurred.

[188] I expand on this point in more detail in the final chapter of this dissertation.

Chapter 5 – Foreknowledge and Moral Responsibility

In the previous two chapters, I grappled with two different traditional incompatibilist viewpoints, what have come to be called "leeway" incompatibilism and "source" incompatibilism. In the second chapter, I broadly defended Frankfurt's strategy for arguing that the ability to do otherwise than we actually do is not a necessary condition for moral responsibility. And in the last chapter, I argued that the most compelling incompatibilist strategy for motivating an "ultimate source" requirement winds up, on closer inspection, falling short of the mark. In this chapter, I would like to turn my attention to a pair of more recent, and somewhat less conventional, arguments in favor of incompatibilism. The two arguments are distinct, but they do share some similarities. They both take the form of direct attacks on compatibilism; they are attempts to show that compatibilism lacks the resources to give the intuitively proper account of the attitude that should be taken towards agents on the basis of future actions of theirs. In that sense, they are both arguments that draw on the perplexing role that *time* plays in our judgments of moral responsibility. Let us turn to the first of these arguments, which being with the curious concept of "prepunishment".

I. What is Prepunishment?

Is it ok to punish people for crimes that they haven't committed yet? Intuitively such a practice seems grossly unjust to say the least. Before a person has committed a crime they are still innocent of the crime, and it is of course immoral to punish the innocent. However some such as Christopher New[189] have argued that this is a baseless temporal bias. Granted, *epistemic* limitations may prevent us from ever actually

[189] Christopher New, "Time and Punishment," *Analysis* 52 (January 1992): 35-40.

punishing people before they commit crimes. But New argues that if we could predict with a reasonable degree of certainty that a person was going to commit a crime in the future, then situations may arise in which it is desirable to prepunish. More recently, Saul Smilansky has tied the question of prepunishment to the free will debate.[190] Smilansky argues that there is a principled way to resist the temptation of prepunishment, but that this strategy assumes the falsity of determinism and hence is open only to the libertarian, not to the compatibilist. Smilansky concludes that compatibilism thus winds up being much more radically revisionist about morality that its proponents would like, thus strengthening the case for incompatibilists who argue that common sense morality cannot be reconciled with determinism.

In what follows, I will offer a response on behalf of the compatibilist. First, I will describe the case for prepunishment (from New). I will then consider Smilansky's argument that prepunishment exhibits a lack of appropriate regard for people as persons, and argue that contra Smilansky that prepunishment is as much an issue for the libertarian as it is for the compatibilist – either Smilansky's favored strategy is available to both, or it is available to neither. And in the final section, I will gesture at some considerations that weigh against prepunishment generally, and may provide a way to resist prepunishment that is somewhat different from what Smilansky suggests and should be open to compatibilists (as well as libertarians).

II. The Case For Prepunishment

New offers an example to illustrate when prepunishment might be acceptable, if not required. Imagine a person, Algy, who intends to and actually is going to speed

---

[190] Saul Smilansky, "Determinism and Prepunishment: The Radical Nature of Compatibilism," *Analysis* 67, no. 4 (2007): 347-49.

tomorrow. Both Algy and the local officer Ben have this knowledge, and they both know that if Ben does not issue a citation for the speeding violation today, before the offense has occurred, Algy will skip the country and never be fined. So Officer Ben issues Algy a ticket the day before the crime, which Algy pays. The next day he goes on to break the speed limit just as described in the citation. Is there anything wrong with what Ben does in this case?

One natural objection that springs to mind is that it is wrong to prepunish Algy because until he commits the crime he is *still innocent*. But New claims that we can distinguish two versions of this basic moral intuition, one of which prohibits prepunishment and one of which allows it. One version is that it is wrong to punish someone for a crime which he *never* commits, and the other is that it is wrong to punish a person for a crime which he has not committed yet, but *intends* to and *actually will* commit. The first version is less controversial and intuitive, but it doesn't prohibit prepunishing Algy, since he DOES commit the crime (in the near future). The latter version prohibits punishing Algy, New says that it is much less obvious that it is correct. And as New points out, there appear to be strong moral considerations *in favor* of punishing Algy before the crime in this case. After all, if we don't prepunish him, he gets to commit the crime and get away with no punishment at all, which ought to be morally repugnant to anyone with any retributivist leanings. New argues that the fundamental intuition here is that there must be some connection between actual guilt and the punishments we inflict. We must be made to pay for our offenses. Whether we happen to pay for them before the crime or after the crime is, New concludes, entirely beside the point.

Another possible objection New considers is that the sort of case under discussion is not actually a case of prepunishment, but an ordinary case of post-punishment. That is, we might suppose that what Algy is being punished for is not his future crime of speeding, but his forming the *intention* to speed. However, this response clearly doesn't apply to the case we are considering. Whatever we may think about punishing a person for their intentions, in the given case Ben writes a citation *for the act of speeding* that Algy will commit the next day. But still we may suppose that New's case for prepunishment implicitly relies on the fact that Algy has already formed the intention to commit a crime. Would the case for prepunishment stand if we removed Algy's intention? We can imagine that Algy KNOWS that he will speed the next day without having yet formed the intention to speed. Or we can modify the case further and suppose that Algy isn't even aware that he we is going to speed tomorrow. Nonetheless Officer Ben knows with certainty that Algy is in fact going to speed tomorrow, and that unless we fine him now, he will skip the country and we will never have the chance to issue a ticket. It seems that here the same considerations in favor of prepunishment apply as in New's original example - that Algy is going to earn the fine, that there is no special reason not to deal out the punishment before the crime rather than afterwards, and that since we can't punish him after the fact, the only way that justice can be served is if we punish him now. So for the moment let us suppose that it doesn't matter whether Algy already *intends* to commit the crime (a point I will return to later). The considerations that New offers in favor of prepunishment seem to stand with or without the criminal already having formed the intention to break the law.

III. Smilansky's Argument Against Prepunishment

Smilansky's objection to prepunishment can be stated simply. He argues on essentially Kantian grounds that we must respect the future criminal AS an autonomous moral agent, specifically as an agent "capable of not committing the offense."[191] Prepunishment, Smilansky argues, violates this basic principle by not giving the agent an opportunity to refrain from carrying out the criminal act in the future. For instance in New's case, Officer Ben is treating Algy as a mere object to be dealt with, rather than as an autonomous moral agent whose autonomy must be respected. Only by giving Algy the *chance* to decide (perhaps at the last moment) that he should do the right thing and drive the speed limit do we *fully* regard Algy a person, in particular a person with the capacity for improving his moral character and doing his duty. Prima facie this seems like a highly plausible objection to New's defense of prepunishment. However I won't spend time here exploring the merits of Smilansky's solution to the prepunishment temptation. Rather, I would like to focus on the implications that Smilansky alleges that this case holds for the free will debate.

Specifically, Smilansky argues that the objection to prepunishment that he has offered is not open to the compatibilist about free will and determinism. To see why, consider New's example again, this time under the assumption that causal determinism is true. According to the standard compatibilist, Algy still may be fully morally responsible for his act of speeding, in the sense that he is blameworthy, and therefore deserving of whatever punishment people can deserve for such a crime. For the compatibilist it makes no difference that Algy's behavior is causally determined. Now suppose that we do have

---

[191] Saul Smilanksy, "The Time to Punish," *Analysis* 54, no. 1 (January 1994): 52.

the means to calculate with complete certainty that Algy is going to break the speed limit the next day. Should we prepunish him? According to Smilansky the compatibilist has no principled way to say no. At any rate, he seems unable to offer Smilansky's objection - that our respect for Algy's moral autonomy demands that we allow him the opportunity to change his mind, because it is already causally determined that Algy is *not* going to change his mind. Therefore the compatibilist seems committed to accepting the practice of prepunishment (in at least in the sort of case under consideration) and this seems like a substantial revision of ordinary morality. Hence the compatibilist's standard claim that determinism makes little or no moral difference is undermined.

Before turning to the compatibilist response, a distinction is in order. We may understand Smilansky in one of two ways. We may understand him as saying that the compatibilist cannot offer the respect-for-the-agent's-autonomy line simply because the agent in fact WILL NOT change his mind. Or, we may understand Smilansky's claim to be that the compatibilist cannot offer this line because the agent is INCAPABLE of changing his mind.

If we read Smilansky the first way, then the problem is not just a problem for compatibilism. There at least two ways in which the libertarian may (in principle) have to deal with possible present truths about what an agent WILL in fact do in the future. First, the most sophisticated forms of libertarianism that are around today allow that much (if not most) of the time, an free agent's behavior is causally determined by his character. For instance, Robert Kane argues that only a very small subset of our actions – what he calls Self Forming Actions - is in fact indeterministic. Kane allows that the rest of our actions may flow deterministically from our characters; "Ultimate Responsibility, or UR,

does not require that we could have done otherwise (AP) for *every* act done of our own free wills."[192] Nonetheless Kane regards these determined actions as ones that are done freely and for which we are responsible, so long as the character they flow from is one that we formed via our properly indeterministic Self Forming Actions. In the above case, we may suppose that being a speed demon is a deeply entrenched part of Algy's character, and that his act of speeding tomorrow is therefore causally determined by his present character. According to a libertarian like Kane, Algy is still responsible for his action. But then since Algy's act of speeding is causally determined, such a libertarian – like the compatibilist – cannot offer Smilansky's defense (on this first reading of it) against prepunishment.

Furthermore, even the more radical libertarian who argues that ONLY indeterministic actions can ever be done freely and responsibly isn't off the hook. Let's assume that determinism is false, and that Algy has this sort of extreme libertarian free will, meaning that *none* of his (free and responsible) actions are causally determined by anything. So when Algy decides to speed, his act not determined by anything that has gone before. And let's assume further an eternalist or "block" theory of time. As I argued in the second chapter, there is nothing incoherent or inconsistent about accepting both of these two claims.[193] There are simply facts about what Algy will do in the future, though what Algy does is still entirely of his own libertarian free will, not *causally* necessitated

---

[192] See Robert Kane in John Martin Fischer et al., *Four Views on Free Will* (Malden, MA: Blackwell Pub., 2007), 14.

[193] Smilansky responds to a related time-travel objection from Helen BeeBee, but as far as I can discern, he offers no reason to think that libertarian free will is incompatible with a block theory of time. See Helen Beebee, "Smilansky's Alleged Refutation of Compatibilism," *Analysis* 68, no. 299 (2008): 258-60, and Saul Smilansky, "More Prepunishment for Compatibilists: A Reply to Beebee," *Analysis* 68, no. 299 (2008): 260-63.

by anything earlier. Now suppose there is a being of some sort (God, a psychic, a time traveler) who tells Officer Bob today that Algy is in fact going to speed tomorrow. Should Bob prepunish Algy? Giving the reading of Smilansky's argument under consideration, the fact that Algy WILL in fact speed makes it pointless to respect his autonomy by giving Algy the opportunity to change his mind. Therefore once again the libertarian seems committed to prepunishing Algy. The problem of prepunishment arises not from determinism itself, but simply from there being accessible facts of the matter about what people will do in the future. It is a problem that both compatibilists and libertarians will have to grapple with, at least in principle.

At this point it might seem that a more plausible way to understand Smilansky is as saying we have to respect Algy's *capacity* to change his mind about speeding, regardless of whether or not there are any facts about what Algy will actually do. This response still would not be open to the compatibilist (we can imagine the argument going), because if determinism is true, then Algy lacks the capacity to do otherwise than what he actually does. But this way of understanding Smilansky's argument would simply beg the question against some prominent versions of compatibilism. Many compatibilists argue that there is a robust sense account of the capacity to do otherwise than we actually do that is consistent with our being causally determined to do as we actually do.[194] Such a compatibilist could agree with Smilansky that prepunishing Algy is wrong on the grounds that it violates respect for Algy's ability or capacity to change his mind before he acts, even given the certainty that Algy will indeed break the law

[194] Of course as we have seen, some compatibilists deny (or are skeptical of) the claim that we could do otherwise than we actually do if determinism is true – I will discuss them in more detail shortly.

tomorrow. To start with the assumption that the compatibilist cannot make sense of respecting Algy's capacity to obey the law is simply to beg the question against this sort of compatibilism; further argumentation is needed.

IV. A Different Response to Prepunishment

If what I have argued so far is correct, then Smilansky has given us no grounds for thinking that compatibilism is on worse footing the libertarianism. Given that the way to cash out respect for Algy's autonomy is in terms of giving him the chance to do otherwise, one of two results follows. Either possible future facts about what he will do undermine the objection to prepunishment (for both the libertarian and the compatibilist), or else they leave the objection untouched (for both the libertarian and the compatibilist). However this still is an ultimately unsatisfying result. For one, there are independent reasons to reject the conditional analysis of possibility (as I discussed earlier). If so, then the compatibilist's ability to resist prepunishment is still questionable. And for another, there are some independent reasons for questioning the second reading that I offered of Smilansky's argument. As Smilansky argues in a later reply to Stephen Kearns, there would seem to be no point in waiting to give someone a chance to do otherwise when we already know for certain that they won't. If that's right, then the second reading is out entirely, and we're left only with the first, which (as I've argued) doesn't provide any support for resisting prepunishment at all.

With these considerations in mind, I would like to suggest a different response to the problem of prepunishment that doesn't rely on an agent like Algy's capacity to do otherwise. I think pointing to those compatibilists who deny that the capacity to do otherwise is a condition of moral responsibility is helpful. As we have seen, this is the

view of several notable contemporary philosophers, including Harry Frankfurt, John Martin Fischer and Mark Ravizza, Nomy Arpaly, and many others, and it is the view that I defend as well. According to this group, our moral responsibility (and also generally our agency, autonomy, etc.) are completely independent of whether or not we actually possess the genuine ability to do otherwise than we actually do. Such compatibilists would generally reject Smilansky's argument that paying proper respect to Algy as a moral agent has anything to do with his ability to do otherwise than he actually does. Still they may agree with Smilansky that treating Algy as an autonomous agent and not a mere object requires not punishing him for crimes that he has not yet committed. In what follows I will very briefly sketch how one might argue for this conclusion.

One manner in which prepunishing might undermine the requirement to respect person's autonomy is that such punishment would (in at least many sorts of cases) be unintelligible or unreasonable to the person being punished. We require that criminals have to have the capacity to understand the nature of and wrongness of their crime in order to be liable. At least in principle, the criminal should be able to understand why he is being punished. This is one reason why the law has special provisions for those who are insane or lack the mental capacity to understand the nature of their actions. Similarly, someone who is being punished for an action that they haven't committed can hardly be expected to find the punishment to be reasonable, even in principle. It's unreasonable because from the practical standpoint of the future-criminal the crime hasn't happened yet, and it's still *up to the future-criminal* whether or not it will occur. And so on similar grounds, one could argue that prepunishment is unjustifiable.

Of course, one might object that this reply only works for some cases of prepunishment. Recall earlier I distinguished between cases in which the criminal intends to commit and knows about her future crime, those in which she just knows about it, and those in which he lacks even knowledge of it. The requirement that the punishment in principle be sensible (in principle) seems pretty clearly to restrict prepunishment against those who know nothing of their future crimes, but it is less clear against those who do have such knowledge, and even less obvious with those who have already formed the intent to commit their future crimes. But even from the practical standpoint of agents who have formed the intention to commit a future action, whether the future action is going to occur is still *up to her* (even if we know in advance that she in fact will not). Thus when we punish such a person for her actions before they have occurred, we do indeed treat her as an object, *not* as a rational moral agent in control of her actions. This of course does not mean that respect requires that we do *nothing* while we wait around for what we know will happen. If we can *prevent* the crime, then by all means we should. But prevention is a separate matter from punishment. The claim I am defending here is only that in *punishing* a future criminal before she has committed the crime do we fail to properly respect her.

In addition, there is one further point that can be made to argue that prepunishment is unacceptable, also without assuming anything about an agent's capacity to do otherwise, simply by reflecting for a moment on the nature and purpose of punishment. The point is that by punishing people before they have committed crimes, we seem to be giving them license to commit the crimes. It is commonly said that by enduring punishment, criminals pay a debt they owe to society. But if the debt is paid

before the crime has occurred, then it seems that society now owes them something – the right to commit the crime. We can imagine a person like Algy happily paying the fine so that he is free to speed tomorrow. Aside from the absurdity of saying that Algy now has the right to break the law, this plainly undermines one of the fundamental purposes (and justifications) of punishment – that punishment ought deter future criminals. If anything prepunishment instead seems more likely to *encourage* future crimes than discourage them, and this provides further grounds for the compatibilist (and for anyone at all) to resist prepunishment.

If my above arguments hold, then we can resist prepunishment even in the case where Algy knows he will and intends to speed in the future. And we can resist it not by appealing, as Smilansky suggests, to the fact that we have to wait and give Algy a chance to do otherwise (even if we know he won't). Rather, as I have suggested, we can resist prepunishment on the grounds that it is incongruent with the nature of punishment - it fails to do one of the things that punishment is supposed to do punishment (discourage future crimes – in fact if anything, it seems to accomplish the opposite), and on the grounds that it violates the basic respect we owe to people, which in part includes a requirement not to dole out punishments which we should not in principle expect the recipient to find reasonable or intelligible. But even if there are some disanalogies that weaken the argument in cases where Algy intends to and knows that he will commit his future crime, the requirement that punishments be in principle intelligible at the very least *severely restricts* the class of acceptable prepunishment. It would still eliminate at least the *most* unintuitive cases – the ones where the person has no knowledge of or intention

of committing his future crime. And thus Smilansky's claim prepunishment forces a *radical* revision of our ordinary moral intuitions is undermined.

V. God, Foreknowledge, and Blame

I would now like to turn to a second recent and novel strategy for attacking compatibilism and motivating incompatibilism. Like the prepunishment argument, this argument points to an alleged deficiency of compatibilism to account for our intuitive judgments regarding how an agent should be treated on the basis of her actions when they are known in advance. In this case however, the argument turns on the question of *blame* rather than *punishment*, and this argument also involves the familiar element of manipulation. This new strategy for attacking compatibilism has been developed very recently in work by Patrick Todd.[195] Todd's strategy is a twist in familiar manipulation arguments for incompatibilism. As he notes, most traditional arguments have centered around what "we" (as observers) can say about the moral responsibility of manipulated agents. Can we praise or blame them? But Todd argues that by looking at what the *manipulators* may say about such agents, we can motivate a new sort of argument in favor of incompatibilism.

Todd focuses on the example of a God who has created a deterministic universe. This God is omniscient and omnipotent, and knows with perfect detail how every event in the universe will play out from the moment he creates it. Further, he designs it with purpose, deliberately shaping the initial conditions of the universe to produce every event that he wants to occur in the unfolding of that universe. Now suppose that one of the events that occurs in this universe is that a man named Ernie kills another man named

---

[195] Patrick Todd, "Manipulation and Moral Standing: An Argument for Incompatibilism," *Philosophers' Imprint* 12, no. 7 (March 2012): 1-18.

Jones. Ernie is a perfectly normal human with fully developed capacities of rational deliberation, and it is on the basis of rational deliberation that he decides to kill Jones. He satisfies any standard compatibilist account of the conditions for freedom and responsibility. God knew in advance exactly when and how this would happen, and what's more he has deliberately designed the universe so that this exact causal sequence would come about. May *God* blame Ernie for killing Jones?

Todd maintains that it is deeply counterintuitive to say that God can rightfully blame Jones for his action. And he maintains further that the best explanation for this fact is the truth of incompatibilism. It is easy to see how an incompatibilist would explain this fact - since God has fully determined what Ernie will do, he lacks the kind of freedom required for blameworthiness, and thus he cannot be blamed by *anyone* for his action, including God. A compatibilist, it seems, cannot take this line. Since Ernie is by stipulation rational, self-reflective, responsive to moral reasons, etc. (see the various compatibilist conditions for moral responsibility discussed in the previous chapter), the compatibilist should maintain that Ernie is indeed blameworthy for killing Jones.[196] So how can the compatibilist explain what is intuitively so wrong with God's blaming Ernie for the murder? In what follows, I will explore a number of different possible avenues of compatibilist response.

[196] As we saw in the last chapter, this is exactly what McKenna and Mele say about the structurally similar "zygote" argument. I will assess this claim later in this chapter.

VI. Bite the Bullet!

One thing a compatibilist might do is accept McKenna's advice and work to develop a hard-line reply to this new version of the manipulation argument. That is, the compatibilist could assert that since Ernie is indeed fully moral responsible for murdering Jones (given his satisfaction of the compatibilist conditions for the sort of freedom that would ground such responsibility) it is in fact acceptable for anyone, including God, to blame Ernie for his actions. While this is a way the compatibilist could go, it does not seem like a very attractive option. As Todd points out, it seems that this would be a severe intuitive cost for compatibilism - and it would be one that had previously gone unnoticed. And it is worth noting that the counterintuitive cost of taking a hard-line reply here seems more severe than the hard-line reply that McKenna advocates for difficult cases like Case 1 (as discussed in the previous chapter). I will discuss this option again in more detail shortly, but first let us consider some possible ways that the compatibilist might avoid biting such a costly bullet.

VII. God's Hypocrisy

A second, and more promising, sort of line for the compatibilist to take is to argue that there is something about God's moral standing - as the willful designer of every detail the universe, including Ernie's murder of Jones - that makes him uniquely unable to appropriately blame Ernie, in spite of Ernie's moral responsibility. In other words, while it's true that Ernie is blameworthy for killing Jones, God is simply in no position to blame him. Drawing on the work of G.A. Cohen on the subject of "moral standing",[197]

[197] G. Cohen, "Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists?," *Royal Institute of Philosophy Supplements* 81, no. 58 (2006): 113-36.

Todd considers and ultimately rejects a couple of different possible ways that this compatibilist response might go.

One way the compatibilist might press this line is to argue that God would be engaged in a sort of *hypocrisy* if he blamed Ernie for murdering Jones. God's hypocrisy would follow from the fact that he willfully and deliberately brought about conditions sufficient for ensuring the Ernie would kill Jones. This seems to reveal something about God's attitude towards Ernie's action - that he, for whatever reason, approved of it, or at least viewed it as something that all things considered ought to happen. Given the attitude he has displayed in designing the universe in this way, it would be insincere, or "in bad faith", for God to then turn around and blame Ernie for what he has done - in spite of the fact that Ernie is indeed morally blameworthy.

Todd rejects this possible line of compatibilist response, arguing that the mere fact that God in some sense intends or brings it about that Ernie murders Jones does not imply anything about God's moral approval of the murder. Invoking a sort of reply "familiar from the project of theodicy",[198] Todd argues that God could possibly have good reasons for deliberately creating a world in which Ernie murders Jones without thereby morally approving of the murder. He illustrates this point with an example. Suppose a man, Bob, has had several items stolen from his home recently, and he strongly suspects (with good reason) his friend Fred. Bob decides to invite Fred over, while leaving an expensive item out where he thinks Fred will be apt to see it and steal it, and secretly sets up a video camera to catch Fred in the act. Sure enough, Fred steals the expensive item. As Todd notes, Fred acts exactly as Bob intended and desired for Fred to

[198] Todd, "Manipulation and Moral Standing: An Argument for Incompatibilism", 7.

act. And yet, there is plainly no reason why Bob cannot sincerely and legitimately blame Fred for stealing his property.

Todd's reply to this line of compatibilist reasoning appears compelling, at least initially. Todd seems to have shown that it is indeed possible for someone to legitimately blame another for acting exactly as he or she intends or wants the other to act. Yet on further reflection, it is not exactly clear how this argument - drawing from the example of Bob and Fred - is supposed to apply to the case with God. What makes the example with Bob so intuitively compelling is that there is no obvious alternate way in which he could accomplish his legitimate and worthwhile goal of proving that Fred is a thief. It is precisely because Bob is limited (epistemically, in terms of his ability to affect the world, etc.) that we intuitively judge that he can set up the scenario in which Fred does exactly as he intends without morally approving of Fred's action. This consideration, of course, does not apply to God (who, by definition, is completely without limitation in terms of his knowledge and abilities). If God has the power to do literally *anything*, then what possible excuse there for him resorting to setting things up so that Ernie murders Fred to accomplish his goals, regardless of how worthy those goals may be? Why didn't God elect to accomplish those goals by some other less objectionable means? At first glance, it is difficult to imagine what possible reason there is for God to choose to use Ernie's murder to accomplish his aims, unless he morally approves of (or at least, is morally indifferent to) Ernie's action in a way that precludes authentic blame.

Now of course, a thoughtful traditional theist will be quick to point out that there are some candidates for explanations as to why a morally perfect, omnipotent God would create a world in which events like Ernie's murder of Jones occur. This is undoubtedly

why Todd makes reference to the project of theodicy. And it would certainly be well beyond the scope of this paper to argue that none of these attempted explanations succeed. Happily, for my present purposes only two points need be made.

First, even if it is true that there is some successful theodicy that explains why God would create the world in which Ernie murders Jones, this explanation is *at least* far from obvious from the description of the example. The devil, in this case, is in the *lack* of details. The reason why the Problem of Evil is taken to be a strong argument against the existence of God as traditionally conceived (even typically by those who ultimately conclude that this argument fails) is because it is *not* obvious why a perfectly good, omnipotent God would choose to create such a world. Given this lack of obviousness, the claim that there may be some hypothetical explanation as to why God would willfully create a world containing such evil cannot be used to undermine this compatibilist line of response (the line, again, that what explains the *initial intuition* that God cannot blame Ernie is that such blame seems inauthentic or insincere).

The second, related, point is that if there is indeed some compelling theodicy (whatever it may be) that satisfactorily explains why God would create the world in which Ernie murders Jones, then that very theodicy undermines the initial intuition that God cannot blame Ernie. If such a theodicy exists, then God more closely resembles Bob in Todd's example. Just as Bob set things up with the intention that Fred commit theft because it was the only way he could accomplish his perfectly legitimate goal of proving that Fred was a thief, so too (if there is a successful theodicy) does God set things up so that Ernie murders Jones because it was the *only way* he could accomplish his legitimate goals (what exactly these legitimate goals are will, of course, depend on the details of the

hypothetically successful theodicy). And just as Bob can set things up in that way while genuinely blaming Fred (since he had no alternative way to accomplish his goal), so too might God set up the world in which Ernie murders Jones while genuinely blaming Ernie - because he had no alternative way to set up the world while accomplishing his goals. In other words, if there is indeed a successful theodicy, then the alleged "bullet" the compatibilist has to bite in asserting that God can authentically blame Ernie becomes much more palatable.

In short, Todd has not offered a compelling explanation of why this initially compelling line of response, relying on the appeal to God's hypocrisy in blaming Ernie, is not available to the compatibilist. If things are as they initially seem - if there is no good reason why an omnipotent God would *need* to create the world in which Ernie kills Jones in order to accomplish his aims - then the charge that it would be inauthentic or hypocritical for God to blame Ernie stands, and it stands on grounds that are available to the compatibilist. If, on the contrary, it turns out on further reflection that there IS a good reason why an omnipotent God would find it necessary to create the world in Ernie murders Jones, then it seems that God (just like Bob) can legitimately blame Ernie; the initial argument that there is no plausible way for the compatibilist to hold that God can blame Ernie loses its force.

VIII. God's Involvement

Supposing that the appeal to God's apparent hypocrisy fails as a compatibilist line of response, Todd considers another possible compatibilist response. This response is that whatever God's attitude towards Ernie's action - even if he genuinely, deeply, authentically disapproves of it - he is nonetheless *responsible* for it, and thus he cannot

legitimately blame Ernie. This is because, to put it simply, it would be illegitimate to blame someone else for something that you share responsibility for bringing about.

Todd concedes that God is indeed in some sense responsible for the murder - after all, he deliberately brought about conditions necessary for it to occur. But he denies that God is thereby morally blameworthy for the murder, and argues that therefore there is room for God to legitimately blame Ernie for murdering Jones. Todd offers a few examples to support this point. One simple example involves a man, Diego, who is (for perfectly acceptable reasons) going to deny his friend Carmen's request for a ride home. Diego knows that causal determinism is true, and he knows from past experience that Carmen will become unduly enraged when he turns down Carmen's request. Thus, Diego is knowingly bringing about conditions that are causally sufficient to ensure that Carmen will become unduly enraged. Is Diego thereby to blame for Carmen's unreasonable outrage, and is he thereby restricted from blaming Carmen? It seems very clear that he is not. This shows, Todd argues, that it is possible for someone to knowingly bring about causally sufficient conditions for someone else to behave in a certain way and still blame that person for his or her actions. Thus, the compatibilist cannot rely on this sort of reasoning to explain why God cannot blame Ernie for murdering Jones.

As before, Todd's objection to this possible compatibilist line of reply is initially compelling. But again, I think there are some important disanalogies between a case like Diego's and the hypothetical case of God. In my view, the most important difference is the fact that Ernie is only bringing about a very tiny *subset* of the causal conditions that are causally sufficient to ensure Carmen's inappropriate reaction. He is not at all responsible for the most important contributing causal factor - Carmen's character. The

fact that she is the sort of person to have such an over the top reaction to such a minor thing would be traceable to other factors - Carmen's upbringing, past experiences, past choices, etc. The fact that Diego is clearly not responsible for any of those myriad other significant causal factors explains why he is not blameworthy for Carmen's behavior - and thus it remains plausible to claim that he can blame her for it.

God, of course, cannot avail himself of the same excuse regarding Ernie's murder of Jones. By stipulation, God is directly responsible for *every single causal detail* of the universe, and the combination of all of those factors was causally sufficient to ensure that Ernie killed Jones (as God knew). Since God is responsible for *all* of the causal details that led to the murder, he is blameworthy for it, and thus we have a plausible explanation for why it would not be legitimate for him to blame Ernie - or so it seems to me. If what I'm suggesting is right, then Todd has not provided a compelling reason to reject this compatibilist line of response to the problem of explaining why God cannot blame Ernie.

IX. Can God Blame Ernie?

In what I've written so far, I have argued that appealing to God's "moral standing" remains a promising strategy for explaining the strong intuition that God cannot legitimately blame Ernie for murdering Jones; Todd's objections to the ways in which a compatibilist might press this line are not convincing. But perhaps my arguments in response to Todd are not convincing, or perhaps there are other difficulties with resting the compatibilist reply on God's moral standing. So let us suppose, for the sake of argument, that appeals to God's moral standing cannot explain why he cannot blame Ernie. Is there any other recourse for the compatibilist?

At this point, "biting the bullet" and simply asserting that God CAN in fact blame Ernie for murdering Jones might begin to seem more attractive than it did initially. Todd's own arguments against the appeal to moral standing might very well help the compatibilist make his case. The compatibilist could argue that what explains the strong intuition that God cannot blame Ernie for murdering Jones is that we naturally make some illicit assumptions about his moral standing (that he is blameworthy for the murder, or that his actions show that he approves of it, etc). Insofar as Todd is successful at rebutting these points, he is also successful at undermining the initial intuition that God cannot blame Ernie.

Todd considers this compatibilist argument and quickly dismisses it: "To this compatibilist line, I do not have much to say besides that, to me, this clearly *does* seem to be an additional cost. That is, even if God is not getting a perverse joy out of determining us to do wrong, it is still considerably mysterious how it could be appropriate for him to determine and blame us."[199] Is Todd right - does defeating or deflating concerns about God's moral standing leave the intuition that God cannot blame Ernie untouched? That is not so clear to me. But even if Todd is right, I don't think it follows that the *compatibilist* has a particular burden to bear.

To see this, let's imagine a scenario that takes place in an *indeterministic* world. Agents in this world have libertarian free will (fill in the details as you like from your favorite libertarian account). As I argued in the second chapter, any reasonable conditions on libertarian free will - including leeway conditions - can be met even if there are future facts about what a person will do (so long as these future facts are not causally

[199] Todd, "Manipulation and Moral Standing: An Argument for Incompatibilism", 15-16.

determined by what happened earlier). There is, in my view, no reasonable libertarian concern about future facts. Given this, that means that even in the world of people with libertarian free will, God can know in advance the facts about how people will react to a given scenario (he has what Luis de Molina called “middle knowledge”) - and he can use this knowledge to design a world that unfolds exactly as he chooses.

So now let us take the libertarian version of Ernie, who - of his pure unfettered libertarian free will - murders Jones. But supposed it is also the case that God planned for Ernie to murder Jones. It was essential to God’s overarching plan for the universe (for whatever reason) that Jones wind up murdered, and God saw that in advance that if he put Ernie in the right place at the right time, that Ernie would in fact murder him (and let us suppose further that God has been doing this sort of thing for Ernie’s entire life - putting him in exactly the right situations in the right time, knowing exactly how Ernie will freely choose in each situation, to ensure that Ernie gradually crafts his murderous character through a series of free self-forming actions). When Ernie finally does murder Jones of his own libertarian free will, exactly as God intended, can God then legitimately blame Ernie for murdering Jones?

To me, the intuition that God cannot legitimately blame Ernie in this new version of the case feels exactly as strong as it did in the causally deterministic version of the case. And it seems to me that considerations related to God's moral standing are good candidates for explaining this intuition. But suppose they aren't - suppose Todd is right that such arguments fail. What recourse does the libertarian have? The libertarian cannot appeal to mere incompatibilism to explain why God cannot blame Ernie, because Ernie's actions are not causally determined. So if "moral standing" cannot explain why God

cannot blame Ernie, then the libertarian - just like the compatibilist - appears to be stuck with the conclusion that God CAN actually blame Ernie, contrary to our initial intuitions. In other words, if the compatibilist is indeed stuck with the conclusion that God can legitimately blame Ernie, this is no "additional cost" for compatibilism; the libertarian, in a similar case, will be stuck with the same sort of conclusion.

X. Is Ernie Really Blameworthy?

Now I would like to discuss an alternate possibility - the possibility that Ernie simply is not morally blameworthy (or at least, that his moral blameworthiness is substantially diminished) - and that THIS is why God cannot blame Ernie. Certainly this is the line that an incompatibilist would be inclined to take. However as noted before, Todd is right this isn't an easy line of response for the compatibilist. If a deterministic causal sequence brought about by mindless naturalistic forces (the laws of physics, the initial conditions of the universe) doesn't undermine a person's moral responsibility for his actions, then why should a similar sequence brought about by intentional design do so? As Todd puts it, "the mere fact that the world was once intentionally arranged in this way should be irrelevant to the facts of responsibility."[200] How can a compatibilist explain in a principled, non ad-hoc way why throwing intentionality into the causal sequence somewhere makes such a difference?

Interestingly, there is a passage in Todd's essay that suggests just how this compatibilist line of response might proceed. Later in the paper, when rejecting the idea that the compatibilist can reasonably deny that God is able to authentically blame Ernie, he says the following: "In keeping with our story, suppose you "wake up" to find yourself

[200] Ibid, 3.

in an afterlife, during which time it is somehow made clear that everything you ever did was part of divinely preordained plan. And then God says to you: "You know, what you did on this occasion was really a horrible thing to have done. What's your excuse? How could you?" Isn't there something deeply unsettling about this scenario? Wouldn't you suppose that something had gone completely wrong?"[201]

What is interesting about this passage is that Todd shifts the focus to the perspective of the person being blamed (Ernie, in this story) to motivate the idea that there would be something inappropriate about God blaming someone for an action when he has deliberately and carefully set up the conditions to ensure that the action would occur. In other words, what makes it seem especially egregious for God to blame Ernie for murdering Jones is that Ernie - if he had all of the facts, if he knew about God's involvement - would have good reason to reject the blame. The force of this intuition seems especially clear regarding God's blame; Ernie has good reason to be incredulous if God later tries to admonish Ernie for doing exactly what God intended him to do and endeavored to ensure that he would do.

This point can be generalized. Once Ernie learns that each and every miniscule detail of his life was crafted and planned by an omnipotent and omniscient manipulator, Ernie would likely feel that his character, choices, and actions were no longer truly *his*; his life would no longer be his own. And feelings of guilt or pride that he might have for his achievements and failures would likely be undermined. And furthermore, it seems quite appropriate that he would come to feel this way. As an agent engaged in practical reason - as someone who seeks out not only the best way to accomplish his goals, but

---

[201] Ibid, 16.

also as someone who aims to figure out which sorts of goals are worth seeking - the news that all of his goals and aims can be traced back to the goals and aims of *someone else*, someone who exercised an overwhelming level of control over every detail of his life, would be devastating indeed. God's preferences seem to take explanatory primacy over everything else.

This discussion connects with the discussion of Mele's zygote argument at the end of the last chapter. As I mentioned there, I find the claim that compatibilists are committed to a hard-line reply to cases involving this sort of theological determinism to be less than convincing. When we do as Todd suggests and consider things from the standpoint of Ernie – when we reflect on the emotional impact of this new knowledge for Ernie, and which sorts of judgments Ernie would find it fitting to accept or reject in the light of such knowledge – I think that this point is strengthened. In a case like this, God's actions and God's preferences seem to take primacy over Ernie's own; Ernie is who he is primarily because God wants him to be.

XI. Conclusion

This chapter has focused on a pair of recent, unconventional, and clever arguments meant to show that compatibilist accounts of freedom and moral responsibility are deficient. Both arguments draw on our intuitive reactions to what to agents on the basis of their known future actions (in the first case, on the appropriateness of punishing someone for what we know she will do, and in the second case, on the appropriateness of a manipulator blaming someone for what he has determined that she will do). I demonstrated that both arguments fail, in some ways for similar reasons. In particular, I showed that both arguments fail to provide problems that are unique to compatibilism; in

both instances, I argued that if the arguments manage to force revisions of our common sense moral intuitions, they are forced on libertarians as well as compatibilists. The examination of these arguments has also helped to bolster some of my earlier points; in particular, we found new support for the idea that compatibilists may resist, in a principled way, having to take a hard-line reply to cases involving theological determinism.

In the final chapter, I will reflect on what the contributions of this dissertation have been. In addition, I will develop a detailed account of deep importance of moral responsibility to our lives, defend moral responsibility from the charge that it is on the whole harmful, and in the process show that the dialectical burden lies with anyone who seeks to reject the compatibilist view of free will and moral responsibility.

Chapter 6 – The Importance of Moral Responsibility

At this point, I would like to offer some reflections on what the contributions of my dissertation project have been so far, and add a few further points to tie things together and strengthen my defense of compatibilism. I will argue that moral responsibility is essential to our moral and emotional lives, that it is not harmful in the ways that some skeptics have suggested, and that therefore there should be a default presumption in favor of compatibilism about freedom and responsibility.

I. What Does it Take to Defend Compatibilism?

I would like to begin by noting that the main contributions of my dissertation have, in a sense, been negative in nature. That is to say, I have spent large sections of this dissertation responding to and critiquing the arguments of others, arguing that attempts to demonstrate the inadequacy of the compatibilist view of freedom and moral responsibility have failed. This is especially true of the last couple of chapters, the bulk of which were spent demonstrating the failure of various arguments (some versions of the manipulation argument, and a novel argument involving the concept of prepunishment) to show the inadequacy of compatibilist accounts of concepts like freedom and responsibility and sourcehood.

To say this is not to deny that my dissertation contains some positive components as well. Earlier in the dissertation, I sketched an account of moral responsibility from the standpoint of practical reason involving the moral emotions. I used this account as part of my defense of the claim that moral responsibility does not require access to alternate possibilities. I also utilized it as part of my defense (after expanding on the account by drawing on some of Lehrer's ideas) of the claim that the "manipulation" component of

some manipulation arguments is what does the explanatory work regarding our intuitions about those cases (not the fact that the agents in question are ultimately determined by outside causes).

Still, the positive account I developed is arguably still somewhat sketchy. My main effort here hasn't been to advance a thoroughly robust, detailed, full-fledged account of the conditions of free and responsible agency. There are a couple of different reasons for this. For one, I think that the work that has been by other philosophers in developing and fleshing out highly detailed accounts of freedom has already been exemplary, and the positive resources they provide are largely adequate for my goal of developing responses what I see as the best available arguments for incompatibilism. I remain heavily influenced by (because I prefer to be) and highly indebted to the work of many who have developed accounts of freedom and responsibility, including some whose basic views are close to my own (e.g. Hilary Bok, Keith Lehrer, Harry Frankfurt, John Fischer, Nomy Arpaly, Jay Wallace, to name a few), and some whose basic stances are very opposed to mine (for example Derk Pereboom, Bruce Waller, and Robert Kane). I do think that a framework for looking at freedom that is developed from the standpoint of practical reason in the ways I suggested can help to motivate and clarify some important aspects of these accounts, but I cannot emphasize strongly enough the great credit I owe to the ideas of others in fleshing out the positive aspects of my account.

II. The Dialectical Burden

There is another, somewhat deeper reason that I have not spent the bulk portion of the dissertation working out a detailed, positive account of freedom and responsibility - that is because, in my view, it is not the most important task at hand. My central goal in

this dissertation has been to support the truth of compatibilism - to defend the claim that our most basic common sense concepts of freedom and responsibility (and the practices grounded in those concepts) remain in tact regardless of what the ultimate arcane revelations of physics and metaphysics might be about the truth or falsity of a thesis like causal determinism. In my view, this ought to be the default position. The claim that we free and morally responsible creatures is grounded in our engagement in practical reason (as I have argued), and our view of ourselves and others as responsible agents - as apt targets of criticism and praise and blame - is deeply connected to our relationships with others and our conceptions of ourselves (as Strawson argued, and as was discussed earlier). Freedom and moral responsibility are essential to the possibility of attitudes like love, admiration, and respect, both for others and for ourselves. To abandon the concepts of freedom and moral responsibility is to severely diminish our emotional and moral lives in many ways. For reasons like these, I think there is a strong prima facie presumption in favor of compatibilism - that these attitudes should prevail even if causal determinism turns out to be true. Without very compelling argumentation, we should not accept that so much of our ordinary personal and moral conceptions and practices could hinge on the truth or falsity of this sort of abstract metaphysical claim; we should instead start from the position that such an abstract metaphysical claim would not change things, at least not substantially.

To say this is, of course, not to say that it is *impossible* that compatibilism could turn out to be false. As I argued before, I think that Peter Strawson was mistaken to claim that metaphysical considerations were completely irrelevant to the evaluation of our ordinary interpersonal practices, because the standards involved in those practices *include*

some metaphysical considerations. It could very well turn out to be the case that causal determinism itself *is* relevant to the conceptions involved in our ordinary practices in such a way that it would undermine the presumption in favor of compatibilism. In my view, the best attempts to demonstrate incompatibilism are efforts to prove that this is the case. When Pereboom develops the four-case argument, for example, he is attempting to show that our ordinary concepts of freedom and responsibility (which manifest themselves in our intuitions about the manipulation examples he discusses) are predicated on the assumption that factors outside of our control are not causally sufficient for our actions, a presumption that would be undermined by the truth of causal determinism. A compatibilist needs to respond to these arguments, to show that there are no incompatibilist assumptions built into the standards by which we judge people to be appropriate targets for praise and blame. In this paper, I have aimed to respond to some of the most compelling versions (at least in my view) of these arguments. The framework I developed helped to answer these arguments - and it helps provide a general strategy for assessing other possible incompatibilist arguments as well.

Most incompatibilists would undoubtedly argue that I have misconstrued the nature of the debate, that I am mistaken about where the dialectical burden lies. I think that at least some incompatibilists, those who are skeptics about free will and moral responsibility, can make a plausible case on this point. The way for such a skeptic to advance this point is to argue that in fact our practices involving moral responsibility are NOT that central to our lives - that we could do away with them entirely without any cost (or at least, without any costs to major that they would be significantly disruptive to our lives and relationships). And perhaps such a skeptic would like to argue that in some

ways our lives could be *improved* if we did away with moral responsibility. Indeed, this is exactly what some prominent skeptics have argued recently, most notably Derk Pereboom and Bruce Waller. If they are right that moral responsibility is dispensable and undesirable, then my claims about the dialectical burden in favor of compatibilism would be weakened. So to complete my defense of compatibilism, I would like to respond to this sort of argument, ultimately defending something like the Strawsonian view that moral responsibility is in some significant ways central to and indispensable (or at least very nearly so) to our social and emotional lives, and that therefore the dialectical burden lies with anyone who advocates doing away with it.

In the next few sections, I will consider some various ways that philosophers have argued for the indispensability of freedom and moral responsibility. I will start with some claims that I think have been somewhat misguided, and then move on to some ways of arguing that I think are more compelling.

III. Can Agency Exist Without Freedom and Responsibility?

One common charge against those who adopt positions like hard determinism and hard incompatibilism is that without freedom and responsibility, we would lose something essential to agency itself. This point can be put various ways. If agency itself is undermined - if we can no longer be legitimately seen as any different from as animals or objects, if the lack of freedom means that we are not fully *persons* (as some have claimed), then the presumption in favor of free will would be strong indeed. However, I think that such claims have been overblown, to say the very least. And furthermore, as such claims are typically rooted in misconceptions about the implications of causal determinism, they are not particularly friendly to my compatibilist project.

One way this argument has frequently been put is that it would be impossible for us to deliberate about two options (at least consistently) unless we accept that both options are really metaphysically open to us. I already argued against this claim at length in the second chapter; I won't rehearse those points again here.

A distinct but related way it is sometimes argued that determinism is a threat to freedom and agency is that determinism would imply that we, as agents, never really contribute anything to the world - that it would mean instead that we are minor cogs in a sea of mechanisms, carried along inexorably in a wave of causation. This contributes to what seems to be a common (at least anecdotally) initial reaction to the idea of causal determinism, that it implies a sort of fatalism - that determinism would imply that there is no reason for me to try to either accomplish or avoid anything, because everything is causally determined to happen regardless of what I try to do. This claim is pretty obviously misguided as well, for reasons that have already been explained at length by a number of compatibilists (and also some incompatibilists, like Waller and Pereboom). As long as my efforts and choices are part of the causal sequence - as long as they have an impact on what happens in the world, as long as at least some of what occurs in the future depends on what I do - it makes perfect sense for me to try to accomplish things, regardless of whether or not causal determinism is true. My discussion about deliberation in deterministic conditions in Chapter 3 applies to this point as well.

It might be argued instead that without praiseworthiness and blameworthiness, we remain agents in the sense that we can deliberate and choose effectively, but we lose something special and important about our status as *persons*. I think this claim is more

plausible; I will explore it in more detail later by exploring the role that moral responsibility plays in our emotional and social lives.

IV. Can Moral Obligation Exist Without Freedom and Responsibility?

Another way that the importance of freedom and moral responsibility has been defended is to argue that without them *all* of morality would be lost. For instance, Spinoza writes that our misguided belief in free will is the reason that "the following abstract notions came into being - praise, blame, right, and wrong."[202] Similarly, Peter van Inwagen writes: "I have listened to philosophers who deny the existence of moral responsibility. I cannot take them seriously. I know a philosopher who has written a paper in which he denies the reality of moral responsibility. And yet this same philosopher, when certain of his books were stolen, said, "That was a *shoddy* thing to do!" But no one can consistently say that a certain act was a shoddy thing to do *and* say that its agent was not morally responsible when he performed it."[203] If something like this were true, then certainly it would bolster my claim about the prima facie presumption in favor of compatibilism. However ultimately I think that such claims are also exaggerated. In this section, I will consider one prominent strategy for arguing that morality itself depends on the existence of moral responsibility.

One obvious way that the loss of freedom might be thought to threaten morality itself can be found in Kant's famous assertion that "ought implies can" – that "When the moral law commands that we *ought* to be better men, it follows inevitably that we must

---

[202] Baruch Spinoza, *The Collected Works of Spinoza*, trans. E. M. Curley (Princeton, NJ: Princeton University Press, 1985), Appendix to Part I.
[203] Peter van Inwagen, *An Essay on Free Will*, 207.

be able to be better men."[204] In other words, it seems wrong to claim that someone has a moral obligation to do perform an action if they are in fact unable to perform it. Thus if we lack the freedom to do otherwise than we actually do, then we would only have moral obligations to perform actions in those cases when we were actually going to perform them. But this seems wrong; if claims of the form "Actions of type x are morally obligatory" are to have any validity, it seems they ought to apply to people regardless of whether or not they are going to perform them.[205]

And one might reason further that if claims about what we morally ought to do (or ought not to do) are undermined by the loss of freedom, then so too are claims about moral rightness and wrongness. Haji argues for this claim by invoking what he calls a standard principle of moral obligation, as follows: "S has a moral obligation to perform [not to perform] A if and only if it is morally wrong for S not to perform [to perform] A."[206] This principle has some plausibility; claims about rightness and wrongness seem intimately tied with claims about obligations. But this implies that if we lack any moral obligations, then nothing that anyone does would be morally right or wrong. Haji concludes that if we live in a deterministic world, and if living in a deterministic world precludes freedom, then the most we can say about actions is that they are morally good or bad.

There are various ways that we might resist Haji's conclusion. Pereboom points out that while the sorts of principles that Haji starts with might initially seem to have

---

[204] Immanuel Kant, *Religion within the Limits of Reason Alone*, trans. Theodore Meyer Greene, Hoyt H. Hudson, and John Silber (New York: Harper & Row, 1960), 46.
[205] Haji argues for this point, see Ishtiyaque Haji, "Moral Anchors and Control," *Canadian Journal of Philosophy* 29, no. 2 (June 1999): 175-203.
[206] Haji, "Moral Anchors and Control", 183.

strong intuitive plausibility, this plausibility needs to be weighed against the extraordinary implausibility of their implications. Any principle that implies "that nothing Hitler ever did was wrong"[207] seems in need of either revision or rejection. As Pereboom argues, while one half of the biconditional in Haji's principle is plausible ("If S has a moral obligation to perform [not to perform] A then it is morally wrong for S not to perform [to perform] A"), the other half - "If it is morally wrong for S not to perform [to perform] A, then S has a moral obligation to perform [not to perform] A" is not so obvious. Pereboom suggests a sort of counterexample:

> For example, suppose you say to an animal-abuser, "You ought not to abuse that animal," but then you find out that he has a psychological condition (which he could have done nothing to prevent) that makes animal-abusing irresistible for him, so that he cannot help but abuse the animal. From my point of view, there is an appreciably strong pull to admitting that the "ought" judgment was false, but there is relatively little to denying that abusing the animal is morally wrong.[208]

Pereboom's counterexample seems fairly plausible to me - at least as plausible as the second half of the biconditional in Haji's principle. So I don't think that Haji has given us decisive reason to think that if moral obligations are undermined by determinism, all claims about rightness and wrongness are undermined as well.

And further, it is not so clear to me that causal determinism would even mean that we need to give up on the notion of moral obligation. For starters, there is the fact that it seems, at least prima facie, that we sometimes have conflicting moral obligations. Haji and others regard this as an impossibility. But many others have found the alleged impossibility of conflicting moral obligations to be highly implausible, and count it as a weakness of the Kantian insistence that "ought implies can". As an example, Joseph

---

[207] Pereboom, *Living Without Free Will*, 145.
[208] Ibid, 147.

Margolis cites the Greek tragedy Antigone,[209] in which the title character has an obligation both to bury her brother and to follow the king's law, which prohibits the burial. As Waller notes, "To the Greeks, this seemed an unfortunate situation, though certainly not impossible."[210] Haji does consider this situation impossible, but it is not obvious (at least to me) why this should be so. The claim that we can sometimes have conflicting moral obligations seems to be at least as intuitively plausible as the claim that ought always implies can in every instance.

The intuitive forcefulness of the "ought implies can" principle is further weakened by ordinary, familiar situations in which it seems, at least on the face of it, that people have obligations to perform actions which they are clearly unable to perform. Waller provides one example - suppose you loan me a large sum of money, but then I suffer a severe financial setback, and I am unable to pay you back.[211] Do I have an obligation to give you your money? It seems clearly that I do - in spite of the fact that I am absolutely unable to do so given my current financial situation. This is simply an obligation that now, due to bad circumstances, I am unable to fulfill. I would be making a mistake if I decide that I am now released from of my obligation to repay you. Instead, a more fitting reaction would be for me to regret the fact that I cannot fulfill my obligation to you, and apologize. Or to take another example - suppose I promise to go see my daughter perform in a dance recital. But then suppose I get delayed getting out of work and miss the train. Should I now feel that I am now released of my obligation to watch her perform? Or should I instead feel that I have a moral obligation that I am now, due to

---

[209] Joseph Margolis, "Excerpts from Ishtiyaque Haji's Discussion with Members of the Audience," *The Journal of Ethics* 4, no. 4 (December 2000): 368.
[210] Waller, *Against Moral Responsibility*, 181-82.
[211] Ibid, 184.

poor circumstance, unable to fulfill? To me, the latter seems much more natural. It's unfortunate that circumstances sometimes make it impossible for us to fulfill our obligations (and indeed, this is a reason to be careful when doing things like making promises or borrowing large sums of money - because you take the risk of winding up with a moral obligation that you cannot fulfill). There is no strong reason that I can see to conclude that it is impossible.

With a little revision, I think that we can salvage what is right about the "ought implies can" principle. After all it should be granted that in many instances, the principle does seem very plausible. For example, while it would certainly be a good thing if I could put an end to man-made global warming right now, it would be absurd to say that I have a moral obligation to do so.[212] And it seems clear that the reason it would be absurd to say that I have such an obligation is that I completely lack any ability to put a stop to global warming. To see why the principle applies here and not to other sorts of ordinary cases in which I might, due to circumstances (or causal determinism) be able to fulfill an obligation, we need to get more clear about the exact meaning of the principle. To do this, I think it is helpful to understand the principle from the standpoint of practical reason. From the standpoint of practical reason, moral obligations make good sense. My understanding of what my moral obligations are can factor into my reasoning about what to do. In this way, moral oughts serve a practical function - they inform and guide our reasoning about what we should and shouldn't do, and they can be used to help us inform and guide the reason practical reasoning of others. And it seems that they can perform this function even if it is the case that, in a given situation, there is only one thing that I

[212] Waller raises a similar example involving the absurdity of telling an observer that he ought to stop a doomed 747 from crashing. See Ibid, 186.

can actually do - provided that I have a general *capacity* for performing the sort of action in question. If we recall for a moment the example in Chapter 3, of Sally deliberating about which of two doors to open even though she knows an alien super-scientist has already locked one of the doors in advance, I think the point is clear. Moral oughts can clearly factor into her process of practical deliberation (suppose that one door leads to her daughter's recital, which she has promised to attend), even if she knows that there is only one possible outcome, that whether or not it is possible for her to live up to that obligation has already been determined in advance.

Given that we can understand moral oughts on the basis of the role that they play in practical reasoning, I think that we can make better sense of the "can" in "ought implies can". The "can" that being is referred to makes much better sense if understood in reference to general capacities, rather than the ability to do otherwise than one does in a given situation. This is why it makes no sense to say that I have an obligation to stop global warming - because I have no general capacity to bring about climate change on a global level. This could also explain why Pereboom's animal-abuser lacks a moral obligation, if we assume that he is truly and completely incapable of ever acting on the relevant moral reasons.[213] And this is why it *does* make sense to say that Sally has an obligation to attend her daughter's recital. Even if it is already causally determined that in this instance she will choose not to, Sally still possesses a general capacity to act on the sorts of moral reasons at hand and do things like attend recitals. And even in cases like where I have become financially unable to repay my loan, or where I miss the train

[213] On this point I must admit that I am a little dubious about whether I agree with Pereboom; I merely mean to point out that *if* he is correct in his assessment, the account of obligation I am describing here has the resources to explain why.

preventing me from keeping the promise to my daughter, it makes sense from the practical standpoint to speak of my broken obligations. Admonishment (whether self-directed, or received from others) or regret or shame for breaking my obligations may not be able to change the outcome in *this* instance, but it can serve to help shape and guide my practical reasoning in the future - to remind to be careful about the obligations which I accept in the future, and to take extra steps to be able to ensure that I will be able to fulfill them. And further, since I am the sort of person who generally cares about others and cares about my moral obligations, a feeling of regret and guilt is an appropriate and fitting reaction to my having failed to satisfy my moral obligations in these particular instances.

So ultimately, I don't see any compelling reason to think that we lose talk of moral obligations and moral rightness or wrongness if we lose freedom in the sense required for moral responsibility. And similarly, I don't think that the loss of freedom and moral responsibility means the loss of morality as a whole.[214] However, if we lose freedom in the sense required for moral responsibility, then we do of course lose at least *one* aspect of morality - we lose the ability to justifiably praise and blame others for their actions, and to appropriately judge them as worthy of praise and blame. Some prominent responsibility skeptics (like Pereboom and Waller) regard this as no great loss, but I think they underestimate the centrality and importance of moral responsibility to our lives in various ways. In the next sections, I would like to consider in detail some of the direct costs of the loss of moral responsibility.

---

[214] For a nice discussion of some of the moral perspectives that could survive the demise of belief in freedom, see Michael Slote, "Ethics Without Free Will," *Social Theory and Practice* 16, no. 3 (1990): 369-383.

V. Can Sincere Regret and Apology Exist Without Moral Responsibility?

The centrality and importance of moral responsibility to our lives is most evident when we reflect on various aspects of our personal relationships with others. One important aspect of our relationships with others concerns the ability to apologize to others when we have wronged them or harmed them or violated one of our moral obligations to them. If the loss of moral responsibility meant that we could no longer sincerely or deeply apologize to others, then our relationships with others would be in a significant way diminished. And there seems to be ample prima facie reason to suppose that moral responsibility is a necessary condition for genuine apology. On the face of it, a central component of a genuine apology is admitting moral responsibility for what you have done. If none of us is ever truly morally responsible for anything, then of course this is impossible. As Trudy Govier and Wilhelm Verwoerd put it, "To apologize for an action is to admit that one did it, that it was wrong and harmful to the victim, and that one was responsible for doing it."[215] Similarly, when Kathleen Gill discusses the "cluster of interrelated beliefs, attitudes, emotions, and intentions"[216] involved in a full apology, one of the five key elements she identifies is "an acknowledgment of responsibility for the action".[217] Gill goes on to suggest that without this element of responsibility, saying "I'm sorry" is not truly an apology, but rather more like an expression of compassion or sympathy (like saying "I'm sorry" when hear that a neighbor has developed leukemia).[218] Such expressions of compassion and sympathy are certainly nice, and they definitely

---

[215] Trudy Govier and Wilhelm Verwoerd, "The Promise and Pitfalls of Apology," *Journal of Social Philosophy* 33, no. 1 (2002): 69.
[216] Kathleen Gill, "The Moral Functions of an Apology," *The Philosophical Forum* 31, no. 1 (2000): 12.
[217] Ibid.
[218] Ibid, 13.

have their place, but if *all* of our apologies were reduced to this, then it seems that an essential component of our relationships with others would be missing.

Some skeptics about moral responsibility, of course, disagree. Waller argues that skeptics can make what he calls full "categorical" apologies, "apologies in which the moral responsibility abolitionist honestly acknowledges having done wrong, sincerely regrets the moral flaw in his or her character, resolves to avoid such wrongful acts in the future, and desires to repair or mitigate the harm caused."[219] Relatedly, Pereboom claims that when you have done wrong, you should reject the claim that we are blameworthy, but you can nonetheless "thoroughly regret what you have done"[220] and "resolve not to perform an immoral action of this kind again, and seek out therapeutic procedures to help treat one's character problems."[221]

These claims might initially seem reasonable. But I find at least one aspect of their claims highly dubious - the idea that one could feel *genuine* regret if one truly does not regard oneself as morally responsible for one's actions. It seems to me that if one truly sees oneself as lacking moral responsibility - if one really thinks that it would be inappropriate for anyone to regard you with reactive attitudes like indignation or anger or resentment or the like because of the immoral thing that one has done - then it is hard to see how one could truly be said to feel regret over one's immoral action. One could of course be *sad* about one's immoral action without feeling that one is morally responsible for it - just as one could be sad about the neighbor who contracts leukemia - but this is different from true regret. If this kind of sadness is all that the moral responsibility

---

[219] Waller, *Against Moral Responsibility*, 191.
[220] Pereboom, *Living Without Free Will*, 205.
[221] Ibid.

skeptic can offer when speaking of "regret", then it is at best a very attenuated sense of regret - in my view not nearly deep enough to play the role that regret plays in a sincere apology.

My claim that the kind of regret that one might experience in the absence of acceptance of moral responsibility is not *genuine* regret can be bolstered by adapting something like McKenna's strategy of turning Pereboom's four-case argument back against him. Recall that Pereboom's argument is that we should generalize from cases like Case 1 to Case 4; in his view, there is no morally significant difference between being thoroughly manipulated by an outside agent and being causally determined by the laws of physics and distant earlier states of the universe. And this generally seems to be the view of the incompatibilist about moral responsibility and causal determinism.[222] As Waller puts the point, "why should the shaping by fortuitous contingencies not undercut freedom if the same shaping by planned contingencies does?"[223]

So let's grant for the sake of argument that incompatibilists like Waller and Pereboom and Kane are right, that there is no morally significant difference between a causally determined agent and one who has been manipulated by an outside agent. And then let us ask - to what extent could a manipulated agent truly regret her actions? To make the question more concrete, let's consider a specific example. Imagine a version of my manipulated agent Riley (discussed in earlier chapters) who walks away from a beach and knowingly allows a child to drown, and then later learns that her actions had been directly programmed and controlled by a nefarious neuroscientist named Jesse. As I

---

[222] This includes libertarian incompatibilists; see for example Robert Kane's discussion of Walden 2 in Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), 65-67.
[223] Waller, *Against Moral Responsibility*, 64.

argued before, Riley would be right to believe that she was not blameworthy for walking away and letting the child drown. Could she at the same sincerely regret her action? It seems obvious that she could not. Riley might be extremely sad that the child had drowned, and she might lament the fact that she had been used as a tool by Jesse to bring about the child's death. But insofar as she *truly* regards Jesse's manipulation as completely undercutting her moral responsibility, it is hard to see how she could genuinely regret the action. If this is right - and if incompatibilists are right that there is no morally significant difference between manipulation and ordinary causal determinism - then it is also hard to see how a moral responsibility skeptic can say that it would *ever* be appropriate to experience true regret. The only way that I can see for such a skeptic to avoid this conclusion in the ordinary deterministic case would be to admit that there is a substantial moral difference between manipulation and ordinary causal determinism - but this admission would undermine one of the major incompatibilist strategies for defending their position.

But let's suppose that I am completely mistaken on this point, and that there is no deep conflict between truly believing that one is not even slightly blameworthy for an immoral action and at the same time also truly, deeply, thoroughly regretting that action. Even if we grant this much, it seems to me that the hard incompatibilist lacks the resources to justify genuine apology. An apology that is offered without any acceptance of blame or blameworthiness seems to be less than a full apology *even if* it may contain the elements of regret and acknowledgement of having done something wrong and the resolve to avoid similar mistakes in the future. I think this point is made clear if we again consider Riley as described above, offering an apology for what Jesse has manipulated

her into doing. If Riley's apologetic attitude is characterized as Waller and Pereboom suggest, then it can be described as something like this (even if this is not exactly how she would verbalize her apology): "I truly and deeply regret letting your child drown, and I'll do everything in my power to make sure I never do anything like this again. Nonetheless, I cannot really be blamed, because an evil neuroscientist who had complete control over my preferences and choices manipulated me into letting your child drown. I'm sorry." If this is the attitude Riley is supposed to have when she expresses her sorrow, then it seems very clear that we she is offering is not a true apology. It is much more akin to the expression of sympathy or sadness that one verbalized when one says "I'm sorry" to the neighbor who has leukemia.

If what I have argued in this section is correct, then there is at least one very significant sort of cost that comes with the abolishment of moral responsibility. Now let us consider some others.

VI. Gratitude in the Absence of Moral Responsibility

There are, of course, positive corollaries of reactive moral attitudes like regret and sorrow, for instance attitudes like appreciation and gratitude. It seems that an overwhelming majority of the ink spilled on the topic of moral responsibility and the reactive attitudes has centered on the negative ones, but the positive attitudes have at least as important a role to play in our lives and our relationships, especially our relationships with those we love. In my view, as I will argue, the capacity for sincere appreciation and gratitude are severely diminished in the absence of moral responsibility - and with it, so too would our loving relationships be diminished.

The reason why positive moral attitudes like gratitude are threatened by the demise of moral responsibility is very similar to the reason why regret and sorrow are threatened. The reason is that a central component of such attitudes is the belief in the sort of freedom required for moral responsibility - the belief that the person to whom you grateful is an apt target for praise and blame for his actions. As Galen Strawson writes, "It seems that we very much want people to be proper objects of gratitude, for example. And they cannot be proper objects of gratitude unless they can be truly responsible for what they do."[224] Lucy Allais expresses the point similarly, saying "feeling gratitude towards someone with respect to an action involves seeing the action as flowing from her free choice."[225] Even Pereboom concedes this point to an extent, saying, "Gratitude might well require the supposition that the person to whom one is grateful is morally responsible for an other-regarding act, and therefore hard incompatibilism might well undermine gratitude."[226]

Waller, by contrast, digs in his heels and insists that gratitude remains unaffected by the demise of moral responsibility. Waller describes the example of a friend named Donna, who is good and loyal and kind and generous, who considers it her moral duty to help her friends, and takes great pleasure in doing so. Waller then goes on to say, "My feeling of gratitude is an appropriate response to Donna, just as my resentment is an appropriate response to Matthew when he gratuitously insults me … Neither Mathew nor Donna is morally responsible, and neither justly deserve reward or punishment, but that

---

[224] Galen Strawson, *Freedom and Belief* (Oxford [Oxfordshire: Clarendon Press, 1986), 308.
[225] Lucy Allais, "Dissolving Reactive Attitudes: Forgiving and Understanding," *South African Journal of Philosophy* 27, no. 3 (2008): 179.
[226] Pereboom, *Living Without Free Will*, 201.

fact - and my recognition of it - does not preclude reactive feelings of resentment and gratitude."[227] One point that I want to briefly mention here is that it is more than a little curious for Waller to be defending the appropriateness of attitudes like resentment and gratitude in this way, given that many other philosophers (including me, as I have discussed in the second chapter) understand moral responsibility in terms of the appropriateness of such attitudes. As I mentioned before, and as is evidenced in this passage, Waller closely ties responsibility with punishment and reward (more closely than it should be tied, as I argued earlier, and as others have argued), and as a result some of his disagreement with most compatibilists seems to hinge on a semantic difference.

A bit later though, we see that what Waller characterizes as gratitude is actually fairly close to what Pereboom claims is left once we *abandon* gratitude. Waller offers the example of an affectionate pet dog who comes up to you in a moment of anguish and "quietly rests her head on your arm, licks your hand, and shows sympathy at your distress … Certainly, you do not consider your faithful dog to be morally responsible, but you have no trouble feeling gratitude for your canine friend's genuine affection."[228] Pereboom mentions a similar example when he talks about the sense of thankfulness that might survive the demise of true gratitude, suggesting "one can also be thankful to a pet or a small child for some favor, even if one does not believe that he is morally responsible. Perhaps one can even be thankful for the sun or the rain even if one does not believe that these elements are backed by morally responsible agency."[229]

---

[227] Waller, *Against Moral Responsibility*, 200.
[228] Ibid, 201.
[229] Pereboom, *Living Without Free Will*, 201.

In my view, examples like these highlight just how far removed the attitude of "thankfulness" that we might have towards those we regard as lacking moral responsibility is from genuine gratitude. Certainly we can, as Pereboom suggests, experience joy and thankfulness when someone (or something) who lacks moral responsibility does something nice for us. But I think we want something deeper than this out of our relationships. If the gratitude and appreciation that we can have for our dearest loved ones is diminished to the level of the kinds of emotional reactions that I can have to pets or even blind forces of nature, then surely something very substantial about our personal relationships has been lost.

VII. Love and Freedom

With the considerations of the preceding couple of sections in mind, I think we are in a good position to understand just how it is that moral responsibility is an essential component of close, loving personal relationships. But first, I would like to say a bit about some ways in which free will is *not* obviously necessary for genuine love.

The idea that genuine freedom in the sense required for moral responsibility might be essential to love has been expressed in various ways by a number of different philosophers. For instance, Kane says "There is a *kind* of love we desire from others - parents, children (when they are old enough), spouses, lovers, and friends - whose significance is diminished by the thought that they are determined to love us entirely by instinct or circumstances beyond their control or not entirely up to them".[230] Similarly, P.F. Strawson suggests that the range of emotions we can experience without the moral

[230] Kane, *The Significance of Free Will*, 88.

reactive attitudes "cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other."[231]

But just what is it about the lack of genuine freedom and moral responsibility that is supposed to undermine the kind of love we have for the people we value most? One common way of explaining this idea is to say that the most valuable sort of love is that love which is freely *chosen*. Put this way, the claim may be understood as something like the following: One of the key things that we value in our loving relationships is that people are involved with us because they freely choose to be. Our friends and lovers care for us because they choose to, and they could choose to care for others instead. The fact that they do not do so - the fact that they freely make the choice to love us - is a key part of why we value and love them in a deeper and richer sense than we love pets or inanimate objects. We might conclude, as W.S. Anglin does, that if love is produced by manipulation or coercion, or "even if the sufficient cause of the "love" is not something easily identifiable like button pressing but something more subtly embedded in the causal structure of the world, it still seems that the love is not authentic."[232]

This idea may seem to have some degree of intuitive appeal at first, but it immediately runs into some obvious objections. Pereboom mentions the example of familial love, such as the love between a parent and child.[233] It strikes me as completely implausible to suggest that, for example, there is any exercise of will (free or otherwise) involved in the instantaneous bond of love that forms between a mother and a newborn child. In fact, there would be something inappropriate for such a bond to have to be

---

[231] Strawson, *Freedom and Resentment, and Other Essays*, 10.
[232] W. S. Anglin, *Free Will and the Christian Faith* (Oxford: Clarendon Press, 1990), 20.
[233] Pereboom, *Living Without Free Will*, 202.

mediated by any effort of will on the part of the mother, free or otherwise. If the mother had to actively will herself to love her new child, we would take it as a sign that something was awry. In this sort of instance, completely unwilled, unfree love seems to be an ideal.

Perhaps it could be argued that the notion of freely willed love remains the ideal for other sorts of cases. Perhaps, for example, it is a necessary (or at least desirable) ingredient of romantic love that our lovers are romantically attached to us because they freely choose to be so attached. But again, it seems like this idea misses something. There is after all a long tradition of romanticizing the idea of involuntary, unwilled love. As Arpaly reminds us, there is a sense in which we find it romantic to say, for example, "it had to be you"[234] - to express the fact that there is no possible way I could fail to love you. In saying this one denies any sort of freedom or exercise of will in falling in love, and yet this in no way detracts from its value. On the contrary, the value being expressed is the very lack of exercise of will (free or otherwise) that is involved. As with maternal love for a newborn child, it seems that something would be off if I had to will myself into falling in love with someone. Freely willed romantic love seems forced, a deficiency rather than an ideal.

And it seems that something similar can be said for the kind of love that exists between friends. It might, as before, be suggested that the ideal is that your friends are emotionally attached to you because they deliberately and freely choose that attachment. But again, this claim seems off the mark. I think many of us are familiar (if we are fortunate) with the experience of meeting a new person and being instantly drawn in,

---

[234] Arpaly, *Merit, Meaning, and Human Bondage*, 4.

feeling a ‘click’ and knowing that a wonderful new friendship has formed. No exercise of will (free or otherwise) seems to be involved in this phenomenon, and it does not seem that the new friendship is lacking in any way as a result. Again the contrary seems to be the case; if an effort of will is required in forcing feelings of friendship, this seems to be more a deficiency than an enhancement of the friendship.

VIII. Love and Moral Responsibility

In the preceding section, I argued that it is misguided to think that the reason that genuine freedom (of the sort required for moral responsibility) is important to our loving relationships is due to an ideal of relationships that are chosen by the free exercise of our wills. As a number of familiar examples illustrate, loving relationships of various sorts are very typically, perhaps even ideally, formed without *any* choice or exercise of will. In this section, I will say a bit about how I think it is that freedom, and in particular moral responsibility, ARE essential to loving relationships.

One insight regarding the way moral responsibility seems absolutely essential to our loving relationships can be drawn from the earlier discussion of reactive attitudes like regret, sorrow, gratitude, and appreciation. As I argued, those attitudes are at least severely diminished in the absence of moral responsibility; the corollary attitudes that survive are impoverished approximations at best. This means that without moral responsibility, the range of attitudes that we can legitimately have towards our loved ones would be much more restricted than we ordinarily suppose. Consider positive attitudes like gratitude and appreciation. In any positive loving relationship, people do kind things for one another - they go out of their way to help their loved ones through times of hardship, they give gifts and offer other gestures to demonstrate affection, they do things

to make their loved ones lives easier and more pleasant in both small and significant ways, and they do things to show the other that they are appreciated. Now imagine that the only attitude we can feel in reaction to all of the myriad ways that loved ones enrich our lives is something akin to the thankfulness that one might feel for the sun or the rain or the affection of a pet, as Pereboom suggests. If this is as deep as the "gratitude" in a loving relationship can go, then it begins to seem like loving adult relationships of the sort we value cannot exist in the absence of moral responsibility.

A similar point can be made regarding negative attitudes like regret and sorrow. As anyone who has been in a loving relationship knows, we do, unfortunately, wrong and harm our loved ones sometimes - and we are sometimes wronged and harmed by them. When this happens, ideally we apologize for the wrong or the harm we have done, which includes acknowledging our moral responsibility for our actions and expressing our deep and genuine regret for what we have done. In offering a sincere apology in this way, we also affirm the importance of our loved ones to us, and ideally, we begin healing the damage that was done by the misdeed. But if (as I argued) genuine regret and sorrow are impossible in the absence of moral responsibility, then loving relationships will have lost another essential component. And further, they will have lost a closely related attitude that enriches and deepens loving relationships - that of forgiveness.

There is arguably also another way that the absence of genuine freedom and moral responsibility diminishes our relationships, one that *does* have to do with choice and the exercise of will. It is not the choice of who to love; rather, this point pertains to the choices we make *after* we are in love. In a monogamous romantic relationship, for example, it's essential that you continuously choose to remain faithful. And in general,

we must make choices to spend time with our friends and loved ones, to make efforts to support them, choices that seem necessary to sustaining our relationships in the long term. Such choices are in this way constitutive components of our loving relationships.

Of course, it might be objected that while choice plays this sort of essential role in many kinds of relationships, there is no reason to think that it must be *free* or *responsible* choice. Pereboom defends this sort of claim, saying "it is difficult to see what is to be added by these continuously repeated decisions being freely willed in the sense required for moral responsibility. It might well be desirable for each participant that the other make these decisions. But that the participants should in addition be praiseworthy for these choices seems hardly relevant."[235] In response, I would like to invite the reader to consider what the value of these choices would be if we learned that a loved one had made them simply because they had been programmed by an outside manipulator to make them. Again, in the standard incompatibilist view, there is no morally significant difference between being manipulated and being causally determined. It seems clear to me that such choices, in being robbed of the kind of freedom required for moral responsibility, would lose a great deal of their value - and the value of the loving relationships sustained by such choices would be greatly diminished as a result.

Finally, I would also like to note that it seems that we also value freedom and responsibility in the ones that we love for its own sake. Let's return to the example of parental love. One of the great delights for parents is seeing how children exercise their wills to grow into unique individuals, with preferences and goals and interests and beliefs which can often diverge quite substantially, in very surprising ways, from those of the

---

[235] Pereboom, *Living Without Free Will*, 204.

parents. Parents love their children for who they become, and part of what they love is that they are, at least to some extent, the free and responsible authors of the people they become. This is not to say that parents would not have substantial love for their children if they lacked such freedom; as we know, parents love their children deeply long before they are morally responsible agents. The claim is just that the development and cultivation of freedom and responsibility adds an extra element, that a parent's appreciation for who his child has freely become - especially if the child freely cultivates praiseworthy goals and virtues - can deepen the love even further. And I think that something similar can be said for our romantic lovers and our friends, that our love for them is enhanced by our admiration for the kinds of people that they freely and responsibly are. If so, then this is one further way that our loving relationships would be deeply diminished by the universal loss of moral responsibility.

IX. Is Moral Responsibility Harmful?

A final point that I would like to touch on briefly is the claim that moral responsibility might in some ways be *harmful*, that it may lead to judgments and practices that are detrimental to humanity as a whole, and that we would therefore be better off eliminating it or replacing it with something else. If this were the case, then my claim that the default presumption ought to be in favor of freedom and responsibility would be substantially weakened. I would like to respond to a couple of those arguments, concluding that while there are some legitimate worries - some cautions we should bear in mind when engaging in practices related to holding people praiseworthy and blameworthy for their behavior - that on the whole, moral responsibility is beneficial to us, particularly when understood from a compatibilist perspective.

The idea that belief in freedom and moral responsibility might overall be harmful seems to be relatively rare viewpoint, at least among recent Western thinkers. Even most skeptics about freedom and responsibility seem to lament (at least to some degree) the loss of our ordinary concepts of freedom and responsibility. For instance, Ted Honderich writes, "Suppose you become convinced of the truth of our theory of determinism. Becoming really convinced will not be easy, for several reasons. But try now to imagine a day when you do come to believe determinism fully. What would the upshot be? It would almost certainly be dismay."[236] And Smilansky argues that even though the notion of libertarian free will is incoherent, it is essential that we embrace the illusion - we couldn't manage otherwise. As he writes, "Illusion is a buffer from threats to our self-conceptions and family relationships on the level of the meaning of our lives. If the ultimate perspective is allowed to poison the appreciation of past concern and effort, or the acknowledgment of fault for past deeds or omissions, it is not only our functioning within families which can be harmed, but the very significance of our relationships and the value we achieve for ourselves within them".[237] But a small subset of skeptics, most notably (and unsurprisingly) Derk Pereboom and Bruce Waller, argue that we can in fact *improve* our lives by abandoning moral responsibility.

One way it might be suggested that moral responsibility is harmful is that it limits the degree to which we engage in investigating and learning about the causes of people's actions. The basic worry is that the more we are inclined to judge, the less we are inclined to try to understand. Indeed, it is not uncommon to hear those who have strong views

---

[236] Ted Honderich, *How Free Are You?: The Determinism Problem* (Oxford: Oxford University Press, 1993), 94.
[237] Saul Smilansky, *Free Will and Illusion* (Oxford: Oxford University Press, 2000), 177.

about responsibility, in particular blameworthiness, to express the claim that this is exactly how things *should* be. Waller provides a striking example of this attitude: "As British prime minister, John Major called for harsher criminal justice measures, especially against juveniles: "Society needs to condemn a little more and understand a little less."[238]

I think this is a legitimate worry, and an important cautionary point for anyone who wants to defend the concept of moral responsibility and the practices that are related to it. We should be aware of and guard against the tendency to be so blinded by our reactive attitudes that we resist or inhibit efforts to understand the causes that drive people to act as they do, especially when they do horrible things. However it must be noted that this is more clearly a problem for those who believe in libertarian freedom and responsibility than it is for compatibilism. Libertarianism implies that there are absolute limits on the degree to which we can understand the causes of actions, because from the libertarian perspective, it is *ultimately* up to us how we act, regardless of the causal factors in our background. It is easy to see how perspective this could lead to an insensitivity and indifference to the undeniable role played by social and cultural factors in producing criminal behavior. A compatibilist, by contrast, can (at least in principle) *fully* acknowledge and seek to understand in complete detail the causal factors that produce criminal behavior, while at the same time regarding individual criminals (so long as they satisfy the relevant compatibilist conditions) as morally responsible for their behavior.

---

[238] Waller, *Against Moral Responsibility*, 283.

A related worry is the degree to which belief in moral responsibility might contribute to an abundance of unhealthy and harmful negative moral emotions like anger. Pereboom discusses some of the harmful effects of moral anger: "Frequently, expressions of moral anger are intended to cause physical or emotional pain. Partly as a result of these problems, moral anger often has a tendency to damage or destroy relationships. In extreme cases, it can provide motivation to take very harmful and even lethal action against another."[239] I think there is something to this worry as well; in excess, there is no doubt that moral anger can cause and has caused great harm to individuals and societies. It has been used to justify every manner of atrocity, and seems to be among the driving forces behind many of the worst conflicts and wars in human history.

I think there a couple points that can be made in response to this worry. For one, it also has to be pointed out that at least *some* degree of moral anger is important for our moral lives. As Pereboom acknowledges, it sometimes "motivates us to resist oppression, injustice, and abuse."[240] If we abandon moral anger *entirely*, as moral responsibility skepticism seems to prescribe, then we lose this important aspect of our moral lives. Further, I think that compatibilists have access to at least some of the same resources as skeptics to ameliorate the worst excesses of moral anger. Pereboom notes that "Philosophers in the Stoic tradition have argued that determinism allows for an increased degree of equanimity in the face of the bad things that happen…The central idea of this position is that if determinism is true, then everything that happens can ultimately be attributed to something encompassing…Then, by psychological identification with this

[239] Pereboom, *Living Without Free Will*, 208.
[240] Ibid.

entity, perhaps by taking on its perspective, one can achieve a sort of acceptance of whatever happens."[241]

There is something in this Stoic argument that I find very attractive. I wouldn't advocate going as far as this Stoic viewpoint suggests and striving for *complete* acceptance of *whatever* happens. And it seems that the belief in compatibilist freedom and moral responsibility would block us from going that far. But a compatibilist could at least go some of the way. In recognizing that she is *ultimately* a product of the same grand cosmological history and the same causal laws that produced the person who has wronged and harmed her (a recognition that makes little sense from the libertarian perspective, it should be noted), a compatibilist might at least gain a significant *degree* of acceptance of what has happened. The compatibilist could be moved the thought that 'there but for the grace of God go I', and she could thereby resist the temptation looking at the offender's transgression "as if from a lofty height".[242] This might mitigate, at least to some degree, her moral anger; what might otherwise develop into an intense and harmful craving for protracted and excessive vengeance and retribution against the one who has wronged her might to some significant degree be diminished. And at the same time, the compatibilist could still recognize the ill and malicious will of the person who harmed her, as exhibited by the harmful action, and her judgment of blameworthiness could remain intact.

To be sure, this is a delicate and challenging balancing of perspectives, but to my view that should hardly be surprising. Our moral emotional lives are rich and complex,

---

[241] Ibid.
[242] Martha Nussbaum, "Equity and Mercy," *Philosophy & Public Affairs* 22, no. 2 (1993): 167.

and we shouldn't expect an account of their appropriateness to be simple. In a way, I think it can be said that both libertarians and free will skeptics are guilty of a similar error - they erroneously eschew complexity for simplicity by completely trying to cut one key element of the truth out of the picture (either the truth that we are indeed ultimately products of the same forces that produce everything, or the truth that we can, on the basis of how we freely choose and act, sometimes be apt targets of the moral reactive attitudes). The reality, in my view, is more rich and nuanced than incompatibilism allows.

X. Conclusion

In this concluding chapter, I have argued for two key points. First, I have worked to show that the abolishment of moral responsibility, which skeptics like Derk Pereboom and Bruce Waller advocate, would have profound costs for emotional and moral lives, especially as pertains to our personal relationships with others. I defended the view that without moral responsibility we would lose important moral emotions like sorrow, regret, forgiveness, gratitude, and appreciation, and that this would in many ways undermine the kinds of loving relationships that are such an important source of meaning and satisfaction in our lives. Indeed, as P.F. Strawson originally argued, it is hard to imagine that it would be possible for us to live our lives this way. The second point is that moral responsibility and the practices related to moral responsibility are not as harmful as skeptics like Pereboom and Waller have suggested, at least not when understood from a compatibilist perspective. On the contrary, compatibilists have the resources to accept that our actions are fully caused and can be interested in pursuing a deep understanding of those causes (as we should be, especially when trying to address large social problems), and compatibilists have the resources to mitigate unhealthy, harmful, and

excessive moral anger. And compatibilists alone have the resources to do this while *also* defending the aptness of our intuitive judgments of moral responsibility, along with all of the important attitudes and emotions connected with moral responsibility.

For these reasons, I think that there should be a prima facie assumption in favor of a compatibilist view of free will and moral responsibility. Until very convincing reasons are given for us to do otherwise, we are justified in trusting our traditional intuitions about freedom and responsibility; the dialectical burden lies with anyone who would seek to reject compatibilism. As I stated before, I don't think that this means (as P.F. Strawson suggested), that the compatibilist view wins outright by default. On the contrary, compatibilism could very well turn out to be false in the final analysis. It could be that all of the practices and emotional attitudes that are grounded in moral responsibility are built on an illusion (and then, perhaps, we would take Smilansky's advice and do our best to maintain that illusion). Incompatibilists have offered some seemingly powerful and compelling arguments against compatibilism, and they must be answered. In the earlier chapters of this dissertation, I worked to show that the best arguments for incompatibilism fail; in my view, incompatibilists have been unable to show that "leeway" or "ultimacy" conditions are necessary for freedom in the sense that grounds moral responsibility. If my responses to the best incompatibilist arguments have been successful, then, given what I have argued about the dialectical burden that incompatibilists bear, we now have strong new reasons to be confident that the compatibilist perspective on freedom and responsibility is the right one.

Bibliography

Allais, Lucy. "Dissolving Reactive Attitudes: Forgiving and Understanding." *South African Journal of Philosophy* 27, no. 3 (2008): 179-201.

Anglin, W. S. *Free Will and the Christian Faith*. Oxford: Clarendon Press, 1990.

Anscombe, G. E. M. *Metaphysics and the Philosophy of Mind.* Minneapolis: University of Minnesota Press, 1981.

Arpaly, Nomy. *Merit, Meaning, and Human Bondage: An Essay on Free Will*. Princeton, NJ: Princeton University Press, 2006.

Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press, 2003.

Ayer, A. J. *Philosophical Essays.* London: Macmillan, 1954.

Beebee, Helen. "Smilansky's Alleged Refutation of Compatibilism." *Analysis* 68, no. 299 (2008): 258-60.

Berofsky, Bernard. "Global Control and Freedom." *Philosophical Studies* 131, no. 2 (2006): 419-45.

Bok, Hilary. *Freedom and Responsibility*. Princeton, NJ: Princeton University Press, 1998.

Buss, Sarah, and Lee Overton. *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge, MA: MIT Press, 2002.

Castañeda, Hector-Neri. *Thinking and Doing: The Philosophical Foundations of Institutions*. Dordrecht: D. Reidel, 1975.

Chisholm, Roderick. "Human Freedom and the Self." In *Free Will*, edited by Derk Pereboom, 143-56. Indianapolis: Hackett Publishing Company, 1997.

Clarke, Randolph. "Deliberation and Beliefs About One's Abilities." *Pacific Philosophical Quarterly* 73 (1992): 101-13.

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

Coates, D. Justin, and Neal A. Tognazzini. "The Nature and Ethics of Blame." *Philosophy Compass* 7, no. 3 (March 2012): 197-207.

Coffman, E. J., and Ted A. Warfield. "Deliberation and Metaphysical Freedom." *Midwest Studies in Philosophy* 29, no. 1 (2005): 25-44.

Cohen, G. "Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists?" *Royal Institute of Philosophy Supplements* 81, no. 58 (2006): 113-36.

Coyne, Jerry A. "Why You Don't Really Have Free Will." USATODAY.COM. January 1, 2012. http://usatoday30.usatoday.com/news/opinion/forum/story/2012-01-01/free-will-science-religion/52317624/1".

Darrow, Clarence. *Attorney for the Damned*. Chicago: University of Chicago Press, 2012.

Davidson, Donald. *Essays on Actions and Events*. Oxford: Clarendon Press, 1980.

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

Dennett, Daniel Clement. *Freedom Evolves*. New York: Viking, 2003.

Ekstrom, Laura Waddell. *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Boulder, CO: Westview Press, 2001.

Feinberg, Joel, and Russ Shafer-Landau. *Reason and Responsibility*. Belmont, CA: Wadsworth, 2013.

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

Fischer, John Martin. *Deep Control: Essays on Free Will and Value*. Oxford: Oxford University Press, 2012.

Fischer, John Martin. "Freedom and Foreknowledge." *The Philosophical Review* 92, no. 1 (January 1983): 67-79.

Fischer, John Martin. *The Metaphysics of Free Will: An Essay on Control*. Cambridge, MA: Blackwell, 1994.

Fischer, John Martin. *My Way: Essays on Moral Responsibility*. New York: Oxford University Press, 2006.

Fischer, John Martin, Robert Kane, Derk Pereboom, and Manuel Vargas. *Four Views on Free Will*. Malden, MA: Blackwell Pub., 2007.

Fischer, John. "Recent Work on Moral Responsibility." *Ethics* 110, no. 1 (October 1999): 93-139.

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66, no. 23 (1969): 829-39.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68, no. 1 (January 1971): 5-20.

Frankfurt, Harry. "Three Concepts of Free Action." *Aristotelian Society Proceedings Supplementary* 49:113-25.

Freud, Sigmund. *Civilization and Its Discontents*. New York: W.W. Norton, 1962.

Gill, Kathleen. "The Moral Functions of an Apology." *The Philosophical Forum* 31, no. 1 (2000): 11-27.

Gomberg, Paul. "Free Will as Ultimate Responsibility." *American Philosophical Quarterly* 15, no. 3 (July 1978): 205-11.

Govier, Trudy, and Wilhelm Verwoerd. "The Promise and Pitfalls of Apology." *Journal of Social Philosophy* 33, no. 1 (2002): 67-82.

Haji, Ishtiyaque. *Freedom and Value: Freedom's Influence on Welfare and Worldly Value*. Dordrecht: Springer, 2009.

Haji, Ishtiyaque. "Moral Anchors and Control." *Canadian Journal of Philosophy* 29, no. 2 (June 1999): 175-203.

Haji, Ishtiyaque. *Moral Appraisability: Puzzles, Proposals, and Perplexities*. New York: Oxford University Press, 1998.

Harris, Sam. *Free Will*. New York: Free Press, 2012.

Harris, Sam. *The Moral Landscape*. London: Bantam, 2010.

Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law*. New York: Oxford University Press, 1968.

Honderich, Ted. *How Free Are You?: The Determinism Problem*. Oxford: Oxford University Press, 1993.

Hume, David. *An Enquiry Concerning Human Understanding, 1748*. Oxford: Oxford University Press, 1999.

Hunt, David. "Moral Responsibility and Unavoidable Action." *Philosophical Studies* 97, no. 2 (1997): 195-227.

James, William. *The Varieties of Religious Experience: A Study in Human Nature*. New York: Modern Library, 1936.

James, William. *The Will to Believe and Other Essays in Popular Philosophy, and Human Immortality.* [New York]: Dover Publications, 1956.

Kane, Robert. "Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson." *Philosophy and Phenomenological Research* 60, no. 1 (January 2000): 157-67.

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

Kane, Robert. "Two Kinds of Incompatibilism." *Philosophy and Phenomenological Research*, 2nd ser., 50 (1989): 219-54.

Kant, Immanuel. *Religion within the Limits of Reason Alone*. Translated by Theodore Meyer Greene, Hoyt H. Hudson, and John Silber. New York: Harper & Row, 1960.

Kaufmann, Walter, and Friedrich Wilhelm Nietzsche. *The Portable Nietzsche*. New York: Penguin Books, 1954.

Lehrer, Keith. *Art, Self and Knowledge*. Oxford: Oxford University Press, 2012.

Lehrer, Keith. "'Can' in Theory and Practice: A Possible Worlds Analysis." In *Action Theory*, edited by Myles Brand and Douglas Walton, 241-70. Dordrecht: D. Reidel, 1976.

Lehrer, Keith. "Cans without Ifs." *Analysis* 29 (1968): 29-32.

Levin, Michael. "Compatibilism and Special Relativity." *The Journal of Philosophy* 104, no. 9 (September 2007): 433-63.

Levy, Neil. "Determinist Deliberations." *Dialectica* 60, no. 4 (2006): 453-59.

Lewis, David K. *Counterfactuals*. Cambridge: Harvard University Press, 1973.

Margolis, Joseph. "Excerpts from Ishtiyaque Haji's Discussion with Members of the Audience." *The Journal of Ethics* 4, no. 4 (December 2000): 368-81.

McKenna, Michael. "Compatibilism & Desert: Critical Comments on "Four Views on Free Will"" *Philosophical Studies* 144, no. 1 (2009): 3-13.

McKenna, Michael. "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument." *Philosophy and Phenomenological Research* 77, no. 1 (2008): 142-59.

Mele, Al, and David Robb. *Rescuing Frankfurt-Style Cases* 107, no. 1 (January 1998): 97-112.

Mele, Al. "A Critique of Pereboom's 'four-case Argument' for Incompatibilism." *Analysis* 65, no. 285 (January 2005): 75-80.

Mele, Alfred R. "Manipulation, Compatibilism, and Moral Responsibility." *The Journal of Ethics* 12, no. 3-4 (2008): 263-86.

Nagel, Thomas. *The View from Nowhere*. New York: Oxford University Press, 1986.

New, Christopher. "Time and Punishment." *Analysis* 52 (January 1992): 35-40.

Nietzsche, Friedrich. *Beyond Good and Evil*. Translated by Walter Kaufmann. New York: Random House, 1966.

Nussbaum, Martha. "Equity and Mercy." *Philosophy & Public Affairs* 22, no. 2 (1993): 83-125.

Peacocke, Christopher. *Being Known*. Oxford: Clarendon Press, 1999.

Pereboom, Derk. "Defending Hard Incompatibilism." *Midwest Studies in Philosophy* 29, no. 1 (2005): 228-47.

Pereboom, Derk. "A Hard-line Reply to the Multiple-Case Manipulation Argument." *Philosophy and Phenomenological Research* 77, no. 1 (2008): 160-70.

Pereboom, Derk. *Living without Free Will*. Cambridge, U.K.: Cambridge University Press, 2001.

Schlick, Moritz. *Problems of Ethics.* New York: Prentice Hall, 1939.

Searle, John R. *Rationality in Action*. Cambridge, MA: MIT Press, 2001.

Slote, Michael. "Ethics Without Free Will." *Social Theory and Practice* 16, no. 3 (1990): 369-83.

Slote, Michael. "Understanding Free Will." *The Journal of Philosophy* 77 (1980): 136-51.

Smart, J. J. C. "Free-Will, Praise And Blame." *Mind* LXX, no. 279 (1961): 291-306.

Smilanksy, Saul. "The Time to Punish." *Analysis* 54, no. 1 (January 1994): 50-53.

Smilansky, Saul. "Determinism and Prepunishment: The Radical Nature of Compatibilism." *Analysis* 67, no. 4 (2007): 347-49.

Smilansky, Saul. *Free Will and Illusion*. Oxford: Oxford University Press, 2000.

Smilansky, Saul. "More Prepunishment for Compatibilists: A Reply to Beebee." *Analysis* 68, no. 299 (2008): 260-63.

Spinoza, Baruch. *The Collected Works of Spinoza*. Translated by E. M. Curley. Princeton, NJ: Princeton University Press, 1985.

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

Strawson, Galen. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75, no. 1-2 (1994): 5-24.

Strawson, P. F. *Freedom and Resentment, and Other Essays*. [London]: Methuen [distributed in the USA by Harper & Row, Barnes & Noble Import Division, 1974.

Stump, Eleanore. "Libertarian Freedom and the Principle of Alternative Possibilities." In *Faith, Freedom, and Rationality: Philosophy of Religion Today*, edited by Jeff Jordan and Daniel Howard-Snyder, 73-88. Lanham: Rowman & Littlefield, 1996.

Taylor, Richard. "Deliberation and Foreknowledge." *American Philosophical Quarterly* 1, no. 1 (1964): 73-80.

Taylor, Richard. *Metaphysics*. Englewood Cliffs, NJ: Prentice-Hall, 1963.

Todd, Patrick. "Manipulation and Moral Standing: An Argument for Incompatibilism." *Philosophers' Imprint* 12, no. 7 (March 2012).

Van Inwagen, Peter. *An Essay on Free Will*. Oxford [Oxfordshire: Clarendon Press, 1983.

Van Inwagen, Peter. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27, no. 3 (1975): 185-99.

Vihvelin, Kadri. "Arguments for Incompatibilism." (Stanford Encyclopedia of Philosophy). March 1, 2011. http://plato.stanford.edu/entries/incompatibilism-arguments/.

Vihvelin, Kadri. "Foreknowledge, Frankfurt, and Ability to Do Otherwise: A Reply to Fischer." *Canadian Journal of Philosophy* 38, no. 3 (2008): 343-72.

Vihvelin, Kadri. "Freedom, Foreknowledge, and the Principle of Alternate Possibilities." *Canadian Journal of Philosophy* 30, no. 1 (March 2000): 1-23.

Wallace, R. Jay., Rahul Kumar, and Samuel Richard. Freeman. *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. New York: Oxford University Press, 2011.

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.

Waller, Bruce N. *Against Moral Responsibility*. Cambridge, MA: MIT Press, 2011.

Watson, Gary. *Agency and Answerability: Selected Essays*. Oxford: Clarendon Press, 2004.

Watson, Gary. "Free Agency." *The Journal of Philosophy* 72 (April 1975): 205-20.

Watson, Gary. "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227-48.

Widerker, David. "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities." *Philosophical Review* 104 (1995): 247-61.

Wolf, Susan. "Asymmetrical Freedom." *The Journal of Philosophy* 77 (1980): 157-66.

Zimmerman, Michael J. *An Essay on Moral Responsibility*. Totowa, NJ: Rowman & Littlefield, 1988.