2015

Hard Incompatibilism and Criminal Law

Samantha M. Berthelette
University of Vermont, sberthel@uvm.edu

Hard Incompatibilism and Criminal Law

Samantha M. Berthelette

Introduction

We tend to have strong intuitions when faced with questions of moral responsibility. We have intuitions about whether criminals are morally responsible for their crimes, whether parents are morally responsible for the way they treat their children, whether students are morally responsible for cheating. But it's not clear what these intuitions indicate, or whether our intuitions about these cases are even cogent. In some cases, analyzing our intuitions about moral responsibility can leave us with more questions than answers. For example, what exactly does it take for an agent to be held morally responsible? This is a question which has sparked many arguments and disagreements in philosophical circles. Derk Pereboom has argued for a position called *hard incompatibilism*, which asserts that we are simply not the sort of creatures who can be morally responsible for our actions. In Chapter I, I recapitulate the way Pereboom sets out the argument. To do this, I explain the different possible ways in which the world could be metaphysically situated. First, I discuss the possibility of our world being *indeterministic*. This means that given the past states of the world and the fixed laws of nature, we are not always causally determined to produce particular actions or decisions. However, even if world is indeterministic, I will explain how there does not appear to be a viable way to salvage the sort of free will required for moral responsibility. Next, I will discuss the possibility of our world being *deterministic*. This means that given the past states of the world and the fixed laws of nature, all of our actions and decisions are always causally determined. Using a series of examples Pereboom has created, I will show how living in a deterministic world is also incompatible with moral responsibility. I conclude this chapter by arguing that, at the very least, Pereboom's arguments provide

good reason for us to doubt whether we can justifiably hold each other morally responsible for our actions.

What I am particularly interested in is how our system of criminal law should change if hard incompatibilism is right. That is, how should we treat criminals and prevent crime if we cannot hold people morally responsible for their actions? This is the focus of Chapters II and III. In Chapter II, I explain that the penal system in the United States involves the intentional, state-sanctioned harm of convicted criminals. I also explain how much of our penal policy relies on the idea that criminals are morally responsible for their actions, and thus deserve to be punished if they act in violation of the law. But if criminals cannot be held morally responsible for their actions, then there is a serious problem with these practices. Since criminals do not deserve to be punished, it would appear as though punishing them would be unjustified. I explain that in order to have properly justified crime control practices, we need to either form a different conception of moral responsibility, or we need to change our current crime control practices. In other words, we need to be *revisionists* about either moral responsibility or about punishment. Manuel Vargas has argued that we ought to do the former, and Pereboom has argued that we ought to do the latter.

Vargas suggests that we reconceptualize moral responsibility, but maintain our responsibility-characteristic practices for pragmatic reasons. Pereboom, on the other hand, suggests that we stop holding people morally responsible[1] altogether and instead revise our system of criminal law to account for this change. After laying out their arguments, I will argue that each one is unsatisfactory. I argue that Vargas's approach

[1] As I will explain in Chapter I, Pereboom argues that we should stop holding people morally responsible *in the basic desert sense.* This leaves open the possibility that we can hold people morally responsible in some other, non-basic desert way.

maintains a severely flawed system, and that there is not good reason to do this. One of the primary motivations for his view is to create better moral agents, but there are better and more effective ways of doing this that result in better overall consequences.

Pereboom suggests that we dispose of our notions of moral responsibility altogether, and instead revise our practices so they do not rely on moral responsibility. After investigating some of the most prominent models for justifying systems of crime control, I explain why these previous suggestions do not work. I then explain Pereboom's own suggestion for a system of crime control—the incapacitation approach. This model suggests that we justify incapacitating criminals by invoking the right to protect ourselves as an analogy to quarantine. As I describe at the end of Chapter II, this view is also unsatisfactory for several reasons. I argue that the incapacitation approach, much like the practice of quarantining, is not a sustainable long-term solution. Additionally, it turns a blind eye to the circumstances that actually create criminality. I also argue that the incapacitation theory would be largely ineffective as an actual practice. Because of these objections, we should reject the incapacitation approach. I argue that what we need is a theory of punishment that keeps the important benefits of both Vargas's and Pereboom's views, but that avoids the problems both of those views encounter. In Chapter III, I propose a theory which can accomplish this goal.

The framework for my positive view is motivated by rule consequentialism, a theory that seeks to establish the best rules that aims to produce the best overall consequences for society. My account, which I will call the *rule-consequentialism model (RCM),* appeals to rule consequentialism as a way to justify certain crime control practices. This model, I believe, accomplishes everything we want to accomplish with a

good system of crime control, inflicts minimal (and justifiable) harm, and maintains our intuitive notions of fairness and justice. The RCM justifies the harm produced by crime control because it implements rules based on the expected consequences of implementing those rules. If hard incompatibilism is true, then the best system of crime control will be one that (1) reaches a high epistemic standard of justification, (2) is most effective in accomplishing its goals, and (3) has the best prospects for real-world implementation and stability. Throughout Chapter III, I argue that the RCM meets each of these standards.

I. The Problem with Moral Responsibility

A. Introduction

In *Harry Potter and the Half-Blooded Prince,* the sixth novel of J. K. Rowling's world-renowned series, Katie Bell is placed under the Imperius Curse.[2] This curse works like a sort of mind-control; the victim is under the complete control of the person who casts the spell. Bell was generally a kind and well-meaning person before this incident, but acted much differently after she was cursed. She even tried to assassinate the widely beloved and admired Albus Dumbledore. However, the other characters in the book—and we the readers—did not hold her morally responsible for what she did while under the Imperius Curse. After all, she was not in control of her actions. She had no choice but to comply with her orders. We don't blame Bell for trying to kill Dumbledore; we blame the one who placed her under the spell in the first place.[3]

This fanciful example helps to illustrate some important issues that have real-world consequences. Our intuitions tell us that Bell is not morally responsible for her actions while under the curse. But why is this, exactly? In some ways, it is easy to point out exactly how Bell's situation differs from all of our situations. We are not being controlled by the Imperius Curse, for example. However, it is not so easy make a clear distinction between Bell's case, in which she is not morally responsible, and an ordinary case of action, in which someone is morally responsible. She did not have control over her actions. But what does having control over one's actions actually mean? She only had

---

[2] Rowling (2005).

[3] Technically, we blame Draco Malfoy even though he was not the one who directly cast the Imperius Curse over Bell. He cursed Madame Rosmerta, who was then forced to curse Katie Bell. But both of these victims were ultimately under the control of Malfoy.

one choice: to comply with her order. But if she were given the choice between killing Dumbledore and killing someone else, would she then be morally responsible? These sorts of distinctions are more difficult to answer than they might appear at first glance.

We often have strong intuitions when it comes to moral responsibility. This isn't just the case for fictitious examples like Katie Bell under the Imperius Curse, but also for real-world examples. We have intuitions about whether criminals are morally responsible for their crimes, whether parents are morally responsible for the way they treat their children, and whether students are morally responsible for cheating. But what's not clear is whether these intuitions make any sense or what it is exactly that these intuitions indicate. The purpose of this chapter will be to explore these sorts of issues by recapitulating the way Derk Pereboom sets up the problem of free will and moral responsibility. I will begin with a discussion of moral responsibility, and explain what it takes for an agent to be held morally responsible for her actions. In §§I.C–I.D, I argue that we are not the sorts of creatures who can be held morally responsible for our actions—a view Pereboom calls *hard incompatibilism.* I will do this by describing the different possible ways our world could potentially be metaphysically situated, and explain how each of these options result in the rejection of moral responsibility. In §I.E, I consider some possible objections to hard incompatibilism. After addressing each one, I will conclude that we have good reason to be hard incompatibilists and we should make changes to the way our system of criminal law is set up. Subsequent chapters will be devoted to determining what those changes should be.

B. Moral Responsibility

There is a trend in the free will literature to discuss free will and moral responsibility as a pair. This is because of the strong relationship these two concepts appear to have. In fact, it is widely believed that free will is a necessary condition for moral responsibility.[4] When determining whether someone is morally responsible for an action, it is not uncommon to ask whether the agent committed an act freely and without compulsion. Similarly, when exploring questions of free will, we are typically concerned with the real-world effects our belief in free will has. This means that in most cases, it is not free will *per se* that we are concerned about at all. Instead, the primary reason we care about the problem of free will is because denying free will often implies losing our ability to justifiably hold people morally responsible for their actions.[5] Of course, there are a number of different views about what free will consists in. For Harry Frankfurt, free will is the ability to act according to one's second-order volition (i.e., the desire that the agent wants her will to be).[6] For John Martin Fischer, free will is the ability to be receptive and responsive to reasons.[7] For Carl Ginet, free will requires the agent to have had the ability to do otherwise, given the fixed past and laws of nature.[8] For the purposes of this paper, I will not defend any particular conception of free will. What I do want to specify is that when I refer to free will, I am referring to the sort of free will required to be held morally responsible for one's actions. Any other conception of "free will" will not be relevant for the present discussion. However, it is necessary to clarify what I mean by moral responsibility.

---

[4] See Levy and McKenna (2009).
[5] See Roskies (2006).
[6] Frankfurt (1971).
[7] Fischer (1994).
[8] Ginet (1990).

When we say that someone is morally responsible for an action, we mean that a person *deserves* a particular sort of response. For example, if Tom is morally responsible for saving a child from drowning, then Tom deserves praise for this good act. If Pete is morally responsible for harming a helpless animal, then Pete deserves to be blamed for this bad act. It is important, however, to make a distinction between basic desert and non-basic desert. *Basic desert* is determined solely by the merit of the agent; the agent's actions and mental states are the only considerations in evaluating desert. *Non-basic desert* takes other factors into consideration in addition to (or instead of) the agent's actions and intentions. For example, a theory of non-basic desert could indicate that Pete deserves to be blamed for his bad act because blaming Pete will produce some good consequence.[9] Non-basic desert considers what a fitting response to an agent would be rather than considering the agent's merit alone. However, I believe basic desert is most closely aligned with our everyday notions of desert. When we say that Pete deserves to be blamed for his actions, we are not usually implying anything about the potential consequences of blaming Pete, or making observations about the way the world should be. When we say that Pete deserves to be blamed, we are saying something specifically about Pete. We are evaluating his actions and intentions, and we are attributing a sort of merit to him. For this reason, I will define moral responsibility in terms of basic desert rather than non-basic desert. This means that an agent *A* deserves response *R* only in virtue of *A*'s actions and intentions, rather than other normative considerations. The next

[9] This non-basic conception of desert is distinct from *that which is better to do.* Although we might agree that blaming Pete will produce good consequences (e.g., Pete changing his evil ways and becoming a better person who does good things for society), the *best* action and the *deserved* action are two distinct concepts—even if we understand desert in a non-basic sense. This is because *that which is better to do* is not a response to Pete; it is at best a response to the state of the world. If Pete deserves blame (in the non-basic sense), then blame is the warranted response to Pete in particular.

step in clarifying the concept of moral responsibility is to determine what it takes for an agent to be morally responsible for an action.

Explanations of moral responsibility often fall into one of two camps. The primary difference between these two camps is whether they take the *Principle of Alternative Possibilities* (PAP) to be true or false.

> **PAP**: A person is morally responsible for her actions only if she could have done otherwise.

Leeway theorists are those who believe PAP is true; they take alternative possibilities to be a crucial component of moral responsibility.[10] *Source theorists* are those who believe PAP is false.[11] They argue that the way the action came about is what's important in determining moral responsibility.[12] More specifically, an agent must be the source of her own actions in the right sort of way in order for her to be held morally responsible for her actions. In other words, the possibility of alternative possibilities is irrelevant in determining moral responsibility. Contemporary source views are typically motivated by Harry Frankfurt's objection to PAP. [13] This objection relies on a counterexample in which a person (Jones) had no alternative possibilities, and yet he is still morally responsible for his action. The counterexample goes roughly like this:

> **Shooting**: Jones can either shoot Smith, or Jones can refrain from shooting Smith. Black wants Jones to shoot Smith, but Black is concerned that Jones might opt out of shooting Smith. Black would rather that Jones decide to shoot Smith on his own, but as a precaution, Black installs a device in Jones's

[10] For example, Ginet (1990), Berofsky (2002), Campbell (1997), van Inwagen (1983).
[11] For example, Fischer (1994), Frankfurt (1971), McKenna (2001), Pereboom (2001; 2014), Stump (1996).
[12] Sartorio (2011).
[13] See Frankfurt (1969).

> brain (unbeknownst to Jones). If Jones showed any initial sign of opting out, then the device would activate and cause Jones to decide to shoot Smith anyway. As it turned out, however, Jones decided to shoot Smith without any interference from Black's device.[14]

In the case of **Shooting**, it seems as though Jones is morally responsible for shooting Smith; after all, Jones shot Smith freely without any interference. However, the example is also set up so that Jones has no alternative possibilities. Even if Jones started leaning towards not shooting Smith, Black would have caused Jones to decide to shoot Smith anyway. These sorts of examples—Frankfurt-style examples, as they are often referred to—provide a counterexample to PAP. Because of this, it seems as though moral responsibility must depend on something other than the mere presence of alternative possibilities. This is not a problem for source theorists. It is, however, a problem for leeway theorists.

There is disagreement about how much of a problem Frankfurt-style examples pose for leeway theories of moral responsibility. Different leeway theorists have come up with different ways of arguing against Frankfurt's conclusion (that PAP is false), but there is a plausible case to be made which shows that source theorists are correct. However, fully defending a source theory against its leeway competitors would be an entirely different project. For the present project, I am only concerned about whether we can have moral responsibility if PAP is false. For the sake of this goal, I will assume that the case for source theories has been sufficiently made. That is, I will assume Frankfurt-style examples show that PAP is false. I will assume that having alternative possibilities

[14] Since his 1969 paper, there have been a significant number of revisions and additions to these sorts of counterexamples, both from Frankfurt himself and from others who have adopted Frankfurt-style views. For example, see Mele and Robb (1998), Fischer (2002), Pereboom (1995), and Haji (1998).

is irrelevant to moral responsibility, and agents can only be morally responsible if they are the ultimate sources of their own actions.[15] The rest of my paper can be viewed as a conditional: if source theorists are right (which they very well might be), then the following arguments can be made.

C. Hard Incompatibilism and Indeterminism

Traditionally, free will debates have focused on whether free will is compatible with *causal determinism*. For a working definition, I will assume causal determinism is true if every state of the world is completely pre-determined by (i) all previous states of the world and (ii) the laws of nature. This means that given the fixed laws of nature and the way the world has been so far, there is only one way the world could turn out at any given time. The world was causally determined to be the way that it is. Those who believe free will is compatible with determinism are *compatibilists,* and those who believe free will is not compatible with determinism are *incompatibilists.* The general theory of incompatibilism can be divided into two subgroups: *libertarians* and *hard determinists.* Libertarians believe that we (at least sometimes) have free will, so causal determinism must be false. Hard determinists believe that causal determinism is true, and so we never have free will. Compatibilists, meanwhile, try to offer conditions under which free will would be possible even in a deterministic world. This framework has been the traditional way of discussing free will and moral responsibility.[16] Either causal determinism settles whether free will exists, or else free will and causal determinism are compatible with one another. The answer to this question would then have various

[15] For a thorough argument in favor of source theories, see Pereboom 2014, chapter 1.
[16] Kane (2005).

implications for moral responsibility. However, Derk Pereboom posits a view outside of this traditional framework: *hard incompatibilism*. Like hard determinists, Pereboom believes that we do not have the sort of free will required to be held morally responsible. However, unlike hard determinists, Pereboom remains agnostic about whether determinism is true. According to hard incompatibilism, we do not have free will if determinism is true, and we do not have free will if indeterminism (the opposite of determinism) is true. In this sense, Pereboom's view runs along the same vein as Galen Strawson's view.[17] Both argue that regardless of whether determinism is true, we are not the sorts of agents that can be the ultimate sources of our own actions.

Before delving into the details of hard incompatibilism, there is an initial attractiveness to the view that is worth highlighting: the view does not make any commitment one way or the other about determinism. This feature is attractive because there *is* a correct answer to whether determinism is true, but we don't know what that answer is. Furthermore, we may never know. After all, the world could *appear* deterministic because we are able to predict events with staggering precision, but the world could still actually be indeterministic. As Adina Roskies explains, "predictability is at best a poor cousin to determinism, and one that can betray its familial roots".[18] We could observe *Event x* follow after *Event y* one million times, and perhaps we could even predict that *Event x* would follow every time *Event y* occurred, but that does not necessarily mean *Event y* causes *Event x*. It might be a mistake to attribute causation to a set of phenomena when, in fact, the phenomena are merely correlated. And, it could be possible for *Event x* to not occur after *Event y*, but we just haven't seen it yet. Likewise,

[17] See Strawson (2000, 2004, 2011).
[18] See Roskies (2012) p. 212.

the world could appear to be indeterministic while actually being deterministic. This sort of phenomenon can be illustrated by computer number generators, or even flipping a coin. These activities appear to be random, but are actually determined by complex algorithms or the laws of physics. Because of these empirical difficulties, any view that relies on either determinism being true or determinism being false rests on unstable grounds. This gives hard incompatibilism an advantage.

To see how hard incompatibilism works, I will explore three possible scenarios. First, I will explore the possibility that causal determinism is false. I will explain how even if this were true, we could not have the sort of free will required to be held morally responsible for our actions. Second, I will explore the possibility that determinism is true and free will is incompatible with determinism. By virtue of these restrictions, we would not have the sort of free will required to be held morally responsible for our actions. Third, I will explore the possibility that determinism is true and free will is compatible with determinism. I will use Pereboom's argument to show how this option is not cogent, and thus we still cannot have the sort of free will required to be held morally responsible for our actions. Therefore, no matter what the initial scenario is, we must reject the idea that we are morally responsible in the basic desert sense.

We will begin with the first possibility I mentioned, that causal determinism is false. In order to assess the viability of free will in the sense required for moral responsibility in an indeterministic world, we need to figure out what exactly is causing our actions. There are three options available to us: *event-causal libertarianism*, *non-causal libertarianism*, and *agent-causal libertarianism.* Event-causal libertarianism is the theory that actions are caused solely by other events, and windows of indeterminacy in

action allow for the existence of free will.[19] Non-causal libertarianism, however, posits that no cause is needed for people to make decisions; agents have the ability to determine actions in a way that is distinct from causation.[20] Agent-causal libertarianism suggests that agents have the power to cause their own decisions without being causally determined to do so.[21] So what are the prospects for each of these libertarian views in an indeterministic world?

Pereboom presents an argument against event-causal libertarianism called the "disappearing agent" objection:

> **The disappearing agent objection:** Consider a decision that occurs in a context in which the agent's moral motivations favor that decision, and her prudential motivations favor her refraining from making it, and the strengths of these motivations are in equipoise. On an event-causal libertarian picture, the relevant causal conditions antecedent to the decision, i.e., the occurrence of certain agent-involving events, do not settle whether the decision will occur, but only render the occurrence of the decision about 50% probable. In fact, because no occurrence of antecedent events settles whether the decision will occur, and only antecedent events are causally relevant, *nothing* settles whether the decision will occur. Thus it can't be that the agent or anything about the agent settles whether the decision will occur, and she therefore will lack the control required for basic desert moral responsibility for it.[22]

In order for the agent to be morally responsible, she would need to be involved in her action in a way that enhances her control to the extent that she can settle which decision occurs. According to the disappearing agent objection, event-causal libertarianism entails that people do not have the power to settle whether the decision will occur. If I have a choice between two options, and the antecedent events render each decision equally probable, then there is no way for the event-causal libertarian to settle which decision will occur. The preceding events do not settle my choice because they make each of my

[19] For example: Mele (1995), Kane (1995), Ekstrom (2000), Balaguer (2004).
[20] For example: Ginet (1990, 1996, 2007), Goetz (2000), McCann (1998), Widerker (2006).
[21] For example: Taylor (1966), Chisholm (1964), Clarke (1993), Griffith (2010).
[22] Pereboom (2014), p. 32.

two decisions 50% likely to occur. And since antecedent events are the only causally relevant factors for settling whether something will happen, nothing at all settles which of my decisions will occur. If this is true, then I lack the control required to be held morally responsible for the decision. Because of this, there is simply not a sufficient causal history between the agent and the action. Without this sort of control, the agent is not the source of her actions; without being the source of her actions, the agent does not have free will. Therefore, event-causal libertarianism entails that the agent lacks the sort of free will required for moral responsibility.

Now I turn to non-causal libertarianism. One way of holding a non-causal libertarian position would be to invoke prima facie causal language to express a purportedly non-causal relation. Pereboom highlights one example of a view of this sort held by Carl Ginet. According to Ginet, "it is possible for there to be an action that was uncaused and also such that it was up to the agent at the time of the action whether it would occur."[23] Ginet continues to explain what it means for an action to be such that it was up to the agent: "it was up to me at time T whether that event would occur only if I made it the case that it occurred and it was open to me at T to keep it from occurring."[24] But it's not at all clear what *causing x* would mean if not *making it the case that x*. It seems as though Ginet is simply skirting around the word "causation" without actually changing the substantive meaning of the idea. If this is the case, then Ginet's non-causal libertarian view falls prey to the same objections as event-causal libertarian views—including the disappearing agent objection.

[23] Ginet (2007), p. 245.
[24] *ibid.*

But one could also be a non-causal libertarian by actually excluding all causal and causal-like language when explaining actions and events. However, it is easy to see how this sort of view quickly excludes the possibility of free will in the sense required for moral responsibility. If there is no causation, then there is no control over making things happen. Since there is no control, agents cannot be the source of their actions in the way required for moral responsibility. All actions would merely be the causal result of randomness and luck, and this is certainly no way to retain moral responsibility.

If non-causal and event-causal libertarianism won't do, then the libertarian might instead choose to adopt an agent-causal view. The agent-causal view entails the existence of agents who possess a certain causal power to effect decisions just in virtue of the kinds of substances agents are. The agent herself must be the initial cause of her own actions, not just be involved in a chain of causes and events *involving* herself, the agent. She must also not be determined to perform those actions. According to agent-causal libertarianism, then, my decision to raise my hand is only free if that decision is not reducible to causation among events and I was not causally determined to do so. There are two possible characterizations of this view. However, Pereboom argues that neither of them turn out to be plausible given our best understanding of physics and natural laws.[25] Given quantum indeterminacy, there are two options. Either agents can follow probabilistic laws, or they can diverge radically from probabilistic laws. First of all, it is incredibly unlikely that agent-caused decisions follow physical probabilities. Even if agents did follow these probabilities, then the causal powers of the agent would be indistinguishable from the causal powers of antecedent events. Thus we have no good reason to believe this is true. However, the idea that agents diverge from these

[25] Pereboom (2014), ch. 3.

probabilistic laws does not hold much water either. This is because we simply have no evidence that such divergences occur. Pereboom notes, "Without evidence for the departures from the natural law that this view predicts, we have insufficient reason to accept it."[26] Because of these concerns, we simply do not have good reason to believe agent-causal libertarianism could be true.

D. Hard Incompatibilism and Determinism

We have now explored all of the options available to the libertarian. Even if the world is indeterministic, there does not appear to be a viable way to salvage the sort of free will required for moral responsibility. The next step will be to assess whether we could justifiably maintain moral responsibility if determinism were true. Recall that determinism is true if and only if every actual state in the world is completely determined by (i) all the past states of the world and (ii) the laws of nature. For now, let's assume that determinism is true. Now the question becomes: is free will compatible with determinism? If the answer is no, then the hard incompatibilist's work is done. Determinism is true, and free will is incompatible with determinism, so we do not have the sort of free will required for moral responsibility. This incompatibilist idea is fairly straightforward. The more interesting option to explore is compatibilism. This will take a little more work.

Pereboom gives us good reason to believe that free will is not compatible with determinism. To do this, he presents a four-case manipulation argument.[27] The idea behind this argument is to present a series of cases that involve a type of agential

[26] *Ibid.*, p. 69.
[27] See Pereboom (1995), (2001), and (2014) for a more full account of the manipulation argument.

manipulation, in which the prominent compatibilist conditions for moral responsibility are satisfied. Without making any relevant changes to the conditions of the cases, Pereboom eventually shows that it is possible for an agent not to be morally responsible for an action even if the compatibilist conditions are satisfied. Through these cases, Pereboom shows that causal determinism is really no less threatening to moral responsibility than being manipulated by another person is. And, since the agent is not responsible in the agent-manipulation cases, we should not hold the agent responsible in the case of causal determinism either. Let's begin with Pereboom's first case:

> **Case 1:** A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision.[28]

In Case 1, a team of neuroscientists press a button to manipulate Plum's neural states right before he begins to deliberate about killing White. This manipulation triggers a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in Plum's decision to kill White. Pereboom goes on to explain that in this case, as with all the others, the prominent compatibilist conditions for moral responsibility have been satisfied. For example, Plum's desire to kill White conforms to his second-order volition—the condition for moral responsibility set forth by Frankfurt.[29] This means that Plum's desire to kill White is what Plum *wants* his will to be. And, Plum's process of deliberation is sufficiently receptive and responsive to reasons—the

[28] Pereboom (2014), pp. 76-77.
[29] Frankfurt (1971).

condition required by Fischer.[30] This means that Plum has the capacity to consider various reasons and react according to his best reasons. The case is set up so that the manipulation is not mere hypnosis or brainwashing; the brain manipulation makes all his mental states cohere with the psychological regularities of genuine agency. Even though the compatibilist conditions for moral responsibility have been met, it would just seem incorrect to say Plum is morally responsible for deciding to kill White. The intuitive response is that Plum is not morally responsible. This suggests, then, that the prominent compatibilist conditions are not sufficient for moral responsibility.

Now consider Pereboom's second case:

> **Case 2:** Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic (as in Case 1), and at times strongly so, with the intended consequence that in his current circumstances he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White.[31]

In Case 2, the neuroscientists programmed Plum at the beginning of his life to reason egoistically most of the time. They programmed Plum in such a way that in his current circumstance of deliberating about killing White, he is causally determined to engage in egoistic reasoning and produce the decision to kill White. This case is designed to appear slightly closer to reality, while maintaining the intuition that Plum is not morally responsible for his decision to kill White. After all, the time at which Plum was manipulated should not affect whether Plum is morally responsible or not. Therefore, it would be unreasonable to attribute moral responsibility to Plum in Case 2 but not in Case 1 because there are no responsibility-relevant differences between the cases.

Here is Pereboom's third case:

---

[30] Fisher (1994).
[31] Pereboom (2014), p. 77.

> **Case 3:** Plum is an ordinary human being, except that the training practices of his community causally determined the nature of his deliberative reasoning processes so that they are frequently but not exclusively rationally egoistic (the resulting nature of his deliberative reasoning processes are exactly as they are in Cases 1 and 2). This training was completed before he developed the ability to alter or prevent these practices. Due to the aspect of his character produced by this training, in his present circumstances he is determined to engage in the strongly egoistic reasons-responsive process of deliberation and to have the first and second-order desires that issue in his decision to kill White.[32]

In Case 3, Plum is an ordinary human being who grew up in a community that trains and educates its youth. One effect of that training is that it causally determined Plum to reason rationally egoistically most of the time. The community completes this youth training before individuals (including Plum) develop the ability to prevent or alter these practices. Due to the aspect of Plum's character which was molded by this training, he is causally determined to produce the decision to kill White. This example is looking much more like the real world than Cases 1 and 2 did. Yet, there are still no responsibility-relevant differences between any of these cases.

This is Pereboom's fourth and final case:

> **Case 4:** Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Cases 1-3). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first and second-order desires.[33]

In Case 4, causal determinism is true. Plum is an ordinary human being in this world who has particularly egoistic reasoning processes. He has the same sorts of reasoning and deliberation processes as in Cases 1-3, and he is again causally determined to decide to kill White. This case is an ordinary example of a deterministic world. Yet, Pereboom

[32] Pereboom (2014), p. 78.
[33] Pereboom (2014), p. 79.

argues, there are still no responsibility-relevant differences between Case 4 and any of the other cases he described.

In each case, Plum decides to kill White. Producing that decision conforms with all the normal compatibilist requirements for responsible action. That is, Plum's actions are not out of character; in each case, he has a second-order volition to kill White; he is responsive to reasons; he has the ability to grasp, apply, and regulate moral reasons; he has the ability to reflect on and revise his decisions based on those moral reasons. Pereboom's first case is the most fantastical of the four, and they get increasingly more "normal" (in the sense of being similar to the way the world actually is), ending with a causally deterministic world case. However, Pereboom argues that there is no relevant difference in any two of the four cases that would allow Plum to retain moral responsibility in one case but not in the other. If Plum is not morally responsible in Case 1, then he is not morally responsible in any of the other cases, including in the case of causal determinism. Pereboom's four-case manipulation argument provides reasons to believe that compatibilism is not a viable option. At the very least, it is now up to the compatibilists to show why the manipulation argument is flawed in such a way that would rescue the sort of free will required for moral responsibility.

E. Objections to the Four-Case Manipulation Argument

There are typically two ways compatibilists have responded to manipulation arguments. Michael McKenna provides some useful language to distinguish these two tactics.[34] There are *soft-line replies* to manipulation arguments and there are *hard-line replies* to manipulation arguments. A soft-line reply would attempt to indicate some

[34] McKenna (2008).

aspect of the manipulation cases that fails to adequately capture the proper compatibilist conditions for moral responsibility. However, this line of objection has typically been a nonstarter. This is because the hard incompatibilist can always amend the cases so that they include whatever conditions the compatibilist thinks is missing. For example, one might object to Pereboom's cases (as I have laid them out here) by saying that Plum's reasoning is not consistent with his true character. But Pereboom could easily then amend the cases to stipulate that this sort of egoistic reasoning *is* consistent with his true character.[35] This is why McKenna proposes a hard-line reply to Pereboom's four-case manipulation argument.[36]

The hard-line approach demands that we amend our intuitions about Cases 1-3. McKenna argues that it is not clearly true that Plum is not morally responsible in any of Pereboom's four cases. To reach this conclusion, McKenna first concedes that there are no responsibility-relevant differences between any of Pereboom's four cases. In fact, McKenna implores other compatibilists to help Pereboom (and other manipulation argument advocates) make the cases as relevantly similar as possible. McKenna's next move is to have us look at the cases coming from the opposite direction: start at Case 4 and end with Case 1. When we start with Case 4, where Plum is an ordinary person living in a deterministic world, it is unclear whether Plum is morally responsible for his decision to kill White. At the very least, we cannot assume that Plum is *not* morally responsible. If we carry this intuition about moral responsibility over between cases—as we should, considering there are no responsibility-relevant differences between them—

[35] In fact, Pereboom does include this stipulation. I omitted some of the compatibilist conditions Pereboom includes in his cases because naming each one would not be necessary for my current project. To read the full explanation of his four cases, see Pereboom (2014).
[36] McKenna (2008).

then we will conclude that it is unclear whether Plum is morally responsible in Case 1. By doing this, McKenna uses Pereboom's own reasoning against him. Since there are no responsibility-relevant differences between the cases, then our intuitions should not change between them. Our intuitions, then, will completely depend on which end we start with. This is clearly problematic, because we cannot conclude anything about what our intuitions actually are about these cases. Thus, McKenna argues that Pereboom's conclusion is flawed.

McKenna's hard-line reply is certainly not the end for hard incompatibilism. In fact, Pereboom offers a convincing response to McKenna's objection.[37] Pereboom explains that it does not matter which way we go about looking at each of his four cases. No matter which way you look at them, you should still come to the conclusion that Plum is not morally responsible in Case 4. The general reason behind this is that a person's intuitions should be open to clarifying considerations. That is, we should look at each case as a set of additional information. If we are open to clarifying considerations when looking at these cases, then it won't matter which way we look at them. This is because Case 1 helps to clarify something about Case 4, whereas Case 4 does not help to clarify anything about Case 1. Our strong intuitions about Case 1 provide useful information about how we evaluate moral responsibility. Case 1 clarifies aspects of Case 4 which are unclear. Since we have no strong intuitions about Case 4, Case 4 does not provide useful information about how we evaluate moral responsibility.

Consider the following analogy I have constructed. Mollie plays the violin. She is very good, and has played for many years. She can read music, play the notes she means to play, and produce a high quality of sound. Now, Mollie has never played the viola

[37] Pereboom (2008).

before. In fact, she has never even seen a viola. The only thing she knows about violas is that they are some type of instrument. Suppose I ask Mollie, "can you play the viola?" She is really unsure of the answer to this question. After all, she is not even sure what a viola is. The most reasonable response for her to give would be something like, "I don't think so." But then suppose I tell Mollie that a viola is just like a violin, except violas are tuned differently. I explain that they have the same number of strings and are both played in the same way—violas just have a lower sound. With this new information, Mollie can now reasonably say that she can in fact play the viola. Notice, however, that this reasoning does not work the other way around. After being informed that the viola and the violin are not different in any playing-relevant respects, it would be completely unreasonable for Mollie to then say, "I guess I'm not sure if I can play the violin, either!"

This example helps to illustrate why McKenna's argument does not work. Mollie knows she can play the violin, and once she understands that violins and violas are similar in all playing-relevant respects, then she can reasonably conclude that she can play the viola. It would be unreasonable for her to assert from this information that she isn't sure whether she can play the violin. Similarly, we assume in Case 1 that Plum is not morally responsible. Once we understand that Case 1 and Case 4 are similar in all morally-relevant respects, we can reasonably conclude that Plum is not morally responsible in Case 4. It would be unreasonable to use this information to reject our earlier assumption that Plum is not morally responsible in Case 1. The four-case manipulation argument is not supposed to force us to maintain our intuitions over cases no matter what. On the contrary, the earlier cases help us to get a clearer understanding of the later cases. Since we have strong intuitions about Case 1, we can use that information

to inform our intuitions about Case 4. The fact that we have unclear intuitions about Case 4 does little to change our intuitions about Case 1. Therefore, McKenna's objection does not stand.

F. Conclusion

In this chapter, I have explored all the viable options for preserving the sort of free will required for moral responsibility. If determinism is true, then it appears as though we cannot be held morally responsible for our actions. This was shown through Pereboom's four-case manipulation argument. If indeterminism is true, then either we cannot be held morally responsible for our actions (because of the disappearing agent objection), or else we would need to opt for an agent-causal view that is incompatible with our best theories in physics. There does not seem to be much reason to do the latter, so we should accept the hard incompatibilist's conclusion: regardless of whether determinism is true, we do not have the sort of free will required for moral responsibility.

In §I.B, I explained the relationship between free will and moral responsibility. I explained that in many cases, the reason we care about free will is because of what comes with it: moral responsibility. Because of this, I chose not to defend any particular conception of free will. I am only concerned with the sort of free will that is required for moral responsibility. After stipulating this, I also defended a source view of moral responsibility. Rather than alternative possibilities being the important factor for evaluating moral responsible, source theorists assert that the way the action came about is what's important. Because I did not offer a full defense of this view, the rest of my argument can be seen as a conditional one.

In §I.C, I explained that causal determinism has been the primary challenge to free will. However, I also explained that hard incompatibilism has an advantage in the free will theory arena. This is because causal determinism is either true or it is not, and we may never know the answer to this question. But hard incompatibilism does not rely on one answer or the other. Hard incompatibilism states that we do not have the sort of free will required to be held morally responsible for our actions regardless of whether determinism is true. To illustrate this claim, I spent the remainder of §I.C discussing the possibility of maintaining moral responsibility in an indeterministic world. I concluded that none of the libertarian options are viable, and thus we cannot be held morally responsible if causal determinism is false.

In §I.D, I attempted to answer the question of whether we could maintain moral responsibility in a deterministic world. Pereboom's four-case manipulation argument shows us that we cannot. At the very least, Pereboom's argument provides good reason for us to doubt the viability of compatibilism. I continued to defend this view in §I.E by laying out possible objections to Pereboom and then showing why they do not hold.

So far I have only set the stage for a more pointed conversation about free will and moral responsibility skepticism. The issue I am particularly interested in addressing is what should happen to our system of criminal law if criminals cannot be held morally responsible for their actions. As I will explain in Chapter II, our current system of punishment is largely reliant on responsibility in the basic desert sense. If incompatibilism shows us that we do not have this sort of responsibility, some major revisions may be in order. The next two chapters are devoted to exploring what these revisions should be.

II. Free Will Skepticism and Punishment

A. Introduction

The penal system in the United States involves the intentional, state-sanctioned harm of convicted criminals. If a person is convicted of a crime, the state could take away her freedom (via incarceration), her property (via fines), or even her life (via capital punishment). Not only does the state inflict intentional harm through punishment, but most people would agree that the state is morally justified in inflicting such harm. One dominant theory that explains why the state is justified in punishing criminals is called *retributivism.* There are two different types of retributivism: pure and impure. According to pure retributivism, the state is morally justified in punishing criminals because, and only because, criminals deserve to be punished. Retributivism is a backward-looking theory of punishment, which means that the potential consequences of inflicting punishment are irrelevant. Instead, the only relevant factors for justified punishment are whether, and to what degree, a criminal deserves to be punished.[38] According to impure retributivism, the state is again morally justified in punishing criminals because the criminals deserve to be punished, but that justification can be overridden by other relevant factors or consequences. Both of these versions, however, require moral desert for punishment to be justified. Many criminal justice scholars view retributivism as the strongest ground for maintaining such an institution.[39] But it is also widely assumed that

[38] Under this conception of desert, I want to reject the idea that a person could deserve punishment for something that has not yet happened. Desert itself, then, is backwards-looking: a person can only deserve a certain kind of treatment based on what has already occurred.
[39] Tadros (2011) p. 24.

ordinary people are retributivists.[40] According to David Dolinko, "The retributive theory is arguably the most influential philosophical justification for the institution of criminal punishment in present-day America."[41]

It is clear that punishment often serves other purposes than just punishing those who deserve to be punished. For example, different forms of punishment can also provide benefits such as protecting society, rehabilitating criminals, and deterring future bad acts.[42] There are also some practices in the United States that are incompatible with retributivism, such as restorative justice programs.[43] However, it is difficult to deny that retributivist motivations are at the heart of the American criminal justice system. We punish people whom we believe deserve to be punished, and we refrain from punishing those whom we believe do not deserve to be punished. When people deserve to be punished, their crimes must typically be the product of *mens rea* (i.e., having a guilty mind). But when people are not morally responsible for their actions, we tend to believe that they should not be punished. This often occurs in cases dealing with minors, cases of insanity, and cases of accidents. Since individuals in these circumstances are typically not morally responsible for their actions, they do not deserve to be punished for those actions. In these sorts of cases, then, we refrain from punishing. In general, our system does its best to dole out punishments only to those who deserve it. In order to deserve punishment, we typically believe that a person must have control over her actions: she must have a choice between right and wrong and freely choose to do wrong.[44]

---

[40] Nichols (2013).
[41] Dolinko (1997), p. 507.
[42] Levy (2012), p. 481.
[43] See Tonry (2013).
[44] "Historically, our substantive criminal law is based upon a theory of punishing the vicious will. It postulates a free agent confronted with a choice between doing right and doing wrong and choosing freely to do wrong." Pound (1927).

If hard incompatibilism is true, then people cannot be held morally responsible for their actions in the basic desert sense. Recall from Chapter I that desert is basic when an agent deserves a certain treatment or response just in virtue of her behavior. Without desert, retributivist punishment is unjustified. But even if retributivism is ruled out as a legitimate justification for punishment, this does not mean that any system of crime control will necessarily be unjustified. The problem hard incompatibilism presents is that we are not the sorts of creatures who can be held morally responsible for our actions in the way we normally conceive of moral responsibility, and therefore our current punishment practices are unjustified. Since inflicting harm morally requires sufficient justification, we need to make some sort of change. There are two routes we could take to try to remedy this problem.

In order to have properly justified crime control practices, we need to either form a different conception of moral responsibility, or we need to change our current crime control practices. In other words, we need to be *revisionists* either about moral responsibility or about punishment. According to Manuel Vargas, "revisionist views are those on which the proposed prescriptive account conflicts with the diagnostic account."[45] A diagnostic account is an account that reflects our ordinary commitments about something. A prescriptive account is an account that reflects how we *should* form commitments to something. In this chapter, I will explore potential ways to pursue both revisionist options. First, I will explain Vargas's theory of revisionism about moral responsibility. I will discuss the benefits that the view has, as well as some potential problems with the view. Next, I will explain Pereboom's theory of revisionism about crime control. I will also discuss the benefits and potential problems of Pereboom's view.

[45] Vargas (2013), p. 85.

I will conclude this chapter by sketching a theory of justification for crime control that maintains the benefits of both Vargas's and Pereboom's approaches, but avoids the major problems of each of their views. A detailed account of this theory will be the subject of Chapter III.

B. Vargas's Moral Responsibility Revisionism

Vargas agrees that libertarianism is not plausible, and thus our normal conceptions of moral responsibility are at risk. However, he rejects the idea that we would need to give up on moral responsibility altogether; in other words, he rejects responsibility nihilism. Vargas's reason for rejecting the nihilist conclusion is grounded in the *principle of philosophical conservation.* According to this principle, we only ought to abandon our standing commitments as a last resort. And, when we do abandon our standing commitments, there is pressure to limit the extent of the revision or elimination. This pressure, Vargas explains, comes from two factors. First, there is a worry about the stability of the web of interlocking justification and explanation that tends to develop among our beliefs. If a major revision is made to one of our beliefs, many of our other beliefs could potentially be affected or threatened. This would create a tear in that web that is much larger and much more difficult to mend than the original intended revision. There are of course times when this sort of massive overhaul of beliefs is warranted, but Vargas worries that making larger revisions than necessary would result in abandoning commitments that need not be abandoned.

The second factor that increases pressure to limit the extent of revision or elimination of our standing commitments has to do with our practices and interactions

with one another. Vargas explains that "in cases where the relevant beliefs are intimately connected with practical matters... the costs of belief revision are particularly high because revision disrupts entrenched dispositions of action."[46] This could potentially be a problem because of the practical role beliefs play in our lives. For example, one reason beliefs are useful is that they help us navigate the world. But in order to maintain this usefulness, our beliefs need to have some degree of *de facto* stability. If we are constantly questioning the legitimacy of every one of our beliefs, then they lose that usefulness. Therefore, according to Vargas, we should not revise our beliefs unless there is some special pressure to do so. Otherwise, we should leave our beliefs as they are.

Because of these two factors, Vargas believes we should follow the principle of philosophical conservation. Although we do have good reason to revise our beliefs about moral responsibility, we should not strip down those beliefs completely. Vargas argues that we should revise our beliefs about moral responsibility in a way that deals with the problem of not having libertarian free will but does not eliminate more of our commitments than is necessary. In particular, Vargas wants to change the way we think of moral responsibility, but avoid getting rid of our responsibility-characteristic practices.[47] Before I explain his revised view of moral responsibility, it is important to better understand what it means for an account to be *revisionist.* According to Vargas,

> Whether an account is revisionist with respect to something depends in part on our ordinary commitments about that thing. If, for example, no one was ever really committed to a divine command theory of morality, then a proposal for the non-divine foundations of morality would not automatically count as revisionist. Similarly, for a theory of free will to

[46] *ibid.*, p. 74.
[47] Vargas does not write about punishment or crime control practices in particular. Rather, he discusses the practices involved in holding one another responsible more generally. This is what he refers to as responsibility-characteristic practices.

> be revisionist, it must propose an account that departs from our ordinary commitments about free will.[48]

This means that a revisionist account must conflict with our ordinary commitments about the subject of the account.

A revisionist account of moral responsibility, then, would conflict with our ordinary conceptions of moral responsibility. According to Vargas, many of us have both compatibilist and incompatibilist intuitions in different contexts. However, Vargas asserts that "for a great many of us, there are contexts in which incompatibilist convictions are really, truly, genuinely part of the conceptual furniture with which we find ourselves."[49] Whereas the diagnostic account of moral responsibility entails the need for alternative possibilities, the prescriptive account Vargas offers does not require alternative possibilities. Vargas proposes the following revisionist account of responsible agency:

> An agent S is a responsible agent with respect to considerations of type *M* in circumstances *C* if *S* possesses a suite of basic agential capacities implicated in effective self-directed agency (including, for example, beliefs, desires, intentions, instrumental reasoning, and generally reliable beliefs about the world and the consequences of action) and is also possessed of the relevant capacity for (A) detection of suitable moral considerations *M* in *C* and (B) self-governance with respect to *M* in *C*.[50]

Vargas's prescriptive view, then, is this: an agent is morally responsible in a situation when that agent has the ability to detect moral considerations, and is able to act according to those moral considerations. So, that's what moral responsibility *should be.* Vargas believes this conception gets closer to what the nature of moral responsibility would have to be, given that our typical notion of moral responsibility is mistaken. The next question to ask, then, is what justifies our practices of holding one another responsible.

---

[48] Vargas (2013), p. 74.
[49] *ibid.*, p. 23.
[50] *ibid.*, pp. 213-14.

Rather than attempt to adopt a theory of moral responsibility that most closely aligns with our folk conceptions of moral responsibility, Vargas suggests that we construct a theory of moral responsibility that achieves the goal we really want to achieve when we hold people morally responsible in the first place—making people better moral agents. Specifically, Vargas believes that our practices of holding one another responsible should derive justification from their effects on creatures like us. Our responsibility-characteristic practices are justified through the consequences of having those practices. It is important not to confuse this concept with the idea that each particular act of holding someone responsible is justified because of the individual consequences that result from this act. Instead, Vargas offers a justification for responsibility-characteristic practices in virtue of their overall consequences. He explains that "the justification arises from the group-level effects of justified norms that are ubiquitously internalized by members of the community and regularly put into practice."[51] Rather than justifying each individual instance of a responsibility-characteristic practice according to its individual consequences, Vargas argues that maintaining responsibility-characteristic practices in general will subsequently cause creatures like us to internalize those practices as norms, and the resulting effects of regularly practicing those norms are sufficient justification for maintaining the practices in the first place. He calls this justification of responsibility-characteristic practices the *agency cultivation model.*

The general idea of the agency cultivation model is this: we are justified in our responsibility-characteristic practices because maintaining those practices helps to cultivate better agency. Vargas believes that our practices of holding people morally responsible help to foster an internalization of moral norms. For creatures like us,

[51] Vargas (2013), p. 172.

practices of moral praise will often encourage certain actions, whereas practices of moral blame will often discourage certain actions. Once people internalize moral norms, they are more likely to act according to the norms as a general rule or habit. Straightforwardly, maintaining a set of responsibility-characteristic practices makes creatures like us more likely to act according to the moral rules more often. If what we want from practices of holding people responsible is just to make people better moral agents, then we need these practices to encourage good moral actions and discourage bad ones. This, Vargas explains, is what justifies responsibility-characteristic practices in general: they help foster better moral agents and lead to the internalization of moral norms across society.

C. Problems with the Vargas-Style Approach

My main objection to Vargas's moral responsibility revisionism can be boiled down to this: he gives up too easily. Vargas's argument might be more effective if we had near-perfectly functioning, optimal agency-cultivating responsibility-characteristic practices currently in place. If this were the case, then we would have good reason to maintain our standing commitments as much as possible so our revisions would not greatly disrupt our current practices. We would be more inclined to keep our responsibility practices in place as they are and pursue only minimal revisions to our beliefs. However, there is a significant sector within our responsibility-characteristic practices that is nowhere near perfectly functioning or optimally agent cultivating: our current system of crime control. Regardless of whether other responsibility-characteristic practices are well-functioning or not, the system of crime control in the United States is deeply flawed. From the massive racial disparities to the high levels of recidivism, and

from functional ineffectiveness to economic turmoil, our current system of crime control is in need of major revisions.[52] Since there are not good reasons to avoid changing our practices (in fact, there are good reasons *to* change our practices), a massive overhaul of beliefs is potentially warranted; there is a special pressure to revise our beliefs. When it comes to a faulty belief that perpetuates such a flawed system of practices, our first priority should not be to minimize changes. Rather, we should try to come up with changes that will fix the most problems most effectively.

Consider this analogy involving our beliefs about how the physical world actually is. For over 1500 years, people tended to believe that the earth was the stationary center of the universe. Many civilizations from the Ancient Greeks to early modern Europe assumed that the moon, the stars, and other planets all revolved around the Earth. We now have very good reasons to believe that this geocentric model of the universe is incorrect. It would have been ridiculous to claim that we should not change our views about geocentricism just for the purpose of following the principle of philosophical conservation. The principle of philosophical conservation, then, can be trumped when there is good reason to make substantive changes to our beliefs. I assert that the beliefs and practices concerning our current system of crime control fits into this category. In other words, we have good reason to make substantive changes to our beliefs and practices entailed by our current responsibility-characteristic practices, and in particular, our system of crime control.

How to solve the problems of racial disparities, recidivism, cost, and ineffectiveness within our current system is not my current project, however. These

[52] For a small sample, see Tadros (2011), Robinson and Darley (2015), Ashworth (2000), Mauer (2001), Blumstein (1982), and Freudenberg (2001).

factors merely bolster the claim that our system of crime control should not be subject to the principle of philosophical conservation. The problem I am most concerned with is the one discussed at the beginning of this chapter: we are not the sorts of creatures who can be held morally responsible for our actions in the way we normally conceive of moral responsibility, and therefore our current punishment practices are unjustified. Vargas suggests that we change our conception of moral responsibility rather than change our responsibility-characteristic practices. He offers two main reasons for this. First, Vargas argues that maintaining our responsibility-characteristic practices helps to produce better consequences. Second, Vargas defends the principle of philosophical conservation. I have argued that there is good reason to change our crime control practices, which trumps the principle of philosophical conservation. Therefore, if there is a way to produce better consequences than the consequences Vargas lays out, then we should take that route instead. Being revisionist about our responsibility-characteristic practices, then, might be the better option.

D. Pereboom's Responsibility Practices Revisionism

Pereboom rightly believes that in order for harm to be justified, the justification must meet a high epistemic standard. Since our current system of crime control is in the business of inflicting harm on people, the justification for our system of crime control must reach that high epistemic standard. If the justification does not reach this standard, then we should stop inflicting such harm. Pereboom's thought here is a simple one: any significant harm we inflict needs justification. Given the end, if it is not overwhelmingly probable that the harm is justified, then it is wrong to inflict that harm.

According to Pereboom, hard incompatibilism does not undermine all responsibility-characteristic practices, just as it does not undermine morality by any means. Many of our practices such as praising and blaming are still potentially justifiable even if responsibility nihilism is right. What hard incompatibilism does undermine, however, are the sorts of practices that cause harm on the basis of moral responsibility in the basic desert sense. This is how our system of punishment comes into question. As explained in the introduction of this chapter, retributivism is a major justification for punishment in the United States. Since retributivism hinges on the idea of moral responsibility in the basic desert sense, it is therefore undermined by hard incompatibilism. This means that retributivism fails to meet a high epistemic standard, and the systematic harm inflicted through punishment should be stopped unless it can be justified in another way.

It is clear that we need some sort of functional crime control system in place. But, if we are to implement and maintain the systematic harm of criminals, then we need proper justification for that harm. Our goals, then, are twofold: first, we must find proper justification for the harm inflicted to control crime; second, we must determine the extent to which such harm is justified. Regarding the first of these goals, we must be able to find a justification that is not undermined by hard incompatibilism. But we cannot just seek out a justification for crime control practices that is unaffected by hard incompatibilism; the justification must also meet a high epistemic standard on its own independent grounds. Once we determine what that justification is, the second goal will be to determine the depth of this justification. In other words, we will need to determine how much harm is justified in which circumstances according to the justification we set out.

As Pereboom sees it, there are four standard methods for justifying punishment. First, there is retributivism. We have already seen how retributivism is undermined by hard incompatibilism, so we can discard this justification right away. The second standard justification for punishment is *moral education*. According to the moral education theory of punishment, punishing criminals is justified because it helps teach them the difference between right and wrong, and aims to foster morally virtuous agents. Moral education theorists often offer the analogy of punishing children for their bad acts. When we punish children, we do not often do so just because we think they deserve to be harmed, but rather because we want to teach them lessons about morality. Pereboom explains the moral education theory further:

> Different levels of severity can communicate distinct levels of seriousness of wrongdoing and, and partly in the consequence of this, punishment can convey important moral boundaries... Finally, coercing a child into behaving in accord with morality could serve to acquaint him with the benefits of moral virtue, which he might subsequently come to value for its own sake.[53]

The general idea is that if we punish criminals, then like children, they will learn from their mistakes and eventually become better moral agents.

Although the moral education theory of punishment is not undermined by hard incompatibilism, there are some major flaws with the theory that render this justification for our current practices unsound. First of all, unlike children, criminals typically already know the moral and legal rules of society. Whereas the goal of punishing children is often to *teach* them a lesson, it is usually not the case that criminals are unaware what they did wrong. Criminals usually already know the codes of conduct within the society of which they are a part. In fact, if the person did not know what they did was wrong, that is usually reason *not* to punish her. Additionally, the psychologies of adult criminals are not

[53] Pereboom (2014), p. 161.

nearly as malleable as the psychologies of children, so the moral education of criminals might not even be particularly effective. Even though our current system may not be ideally set up as an effective moral education model, the high recidivism rates in the United States are evidence of this. A proponent of the moral education theory would expect even our current penal system to educate and change the behaviors of criminals; yet, recent studies show that 76.6% of released prisoners were rearrested within five years of their release in 2005.[54] It's not clear that adding more of a focus on moral education in our system would do anything to significantly reduce that number. There is just not enough evidence to show that moral education would be effective. Because of this, moral education would not be a sufficient justification for intentionally inflicting harm on criminals.

The third standard method for justifying punishment is deterrence. According to deterrence theories, punishment is justified because it helps to prevent criminal wrongdoing. Like the moral education theory, this type of theory is one which does not necessarily rely on the notion of moral responsibility in the basic desert sense. Instead, it merely relies on the consequences of punishment for justification. Pereboom describes a classic construction of a deterrence theory:

> The classic deterrence theory is Jeremy Bentham's.[55] In his conception, the state's policy on criminal behavior should aim at maximizing utility, and punishment is legitimately administered if and only if it does so... The most significant pleasure or happiness that results from punishment derives from the security of those who benefit from its capacity to deter.[56]

However, this utilitarian theory also has its issues. First of all, it seems as though this approach would justify punishments that are overly severe by any reasonable person's

[54] Durose *et al.* (2014).
[55] Bentham (1843).
[56] Pereboom (2014), pp. 163-64.

standard. After all, it makes sense that fewer people would break the rules if the consequences for breaking them were exceptionally harsh. This is problematic because a deterrence-based theory could justify unfair and repugnant consequences like life imprisonment for a mere parking violation. Second, deterrence theories justify the framing and punishing of innocent people in some circumstances. If society thinks there is a very high chance of getting caught for committing a crime, then there is less of a chance that people will commit crimes.[57] All this requires is a scapegoat and good secret-keeping practices. This, again, is an unfair and repugnant consequence of the deterrence model. Pereboom asserts that the moral objections to a deterrence theory are enough to rule it out as proper justification for our punishment practices.

The fourth standard justification for punishment invokes the right of self-defense. Pereboom discusses Daniel Farrell's account.[58] According to this theory, we all have the right of *indirect* self-defense and defense of others. This means that we each have a right to threaten an unjust aggressor in order to prevent the aggressor from harming you or someone else. We also have the right to *direct* self defense and defense of others, which involves actually harming an unjust aggressor to prevent her from harming you or someone else. Farrell explains that the state acts as a sort of proxy for each of us, and so is justified in threatening harm and carrying out harm in order to prevent unjust aggressors from aggressing. Our right to self-defense and the defense of others, then, can ground a theory of punishment that does not depend on a conception of moral responsibility in the basic desert sense. However, as you may have guessed, there are independent issues with this theory as well. Although this theory can properly justify

[57] Wright (2010), Kleiman (2009), Antunes and Hunt (1973).
[58] Farrell (1989).

preventative detention, it is not clear whether the self-defense theory can actually justify punishment. This is because self-defense cannot justify punishing a criminal after he has already caused the harm you were trying to prevent. As Pereboom puts it, "you retain the right to protect yourself and others against him, but not by carrying out the threat designed to prevent a harm that has already occurred."[59] If we seek a well-functioning crime control system, then it should be one that has credibility rather than one that is filled with empty threats. It's not clear that the self-defense theory of punishment can give that to us.

Because all four of the standard justifications for punishment are either undermined by hard incompatibilism or simply have significant problems as independent theories, Pereboom argues that none of them can reach the high epistemic standard of justification needed for the harm inflicted through punishment. This is where Pereboom's own theory of punishment comes in: the incapacitation theory. He suggests that we justify punishing criminals by invoking the right to protect ourselves using an analogy to quarantine. We are justified in incapacitating criminals to protect ourselves and others just as we are justified in quarantining carriers of severe communicable diseases to protect ourselves and others. Even though the criminal may not be morally responsible in a basic desert sense for her criminal actions, we are still justified in incapacitating her in order to protect society from further harm. Neither the criminal nor the disease carrier is morally responsible in this way, but we are still justified in removing them from society. Pereboom explains,

> The core idea is that the right to harm in self-defense and defense of others justifies incapacitating the criminally dangerous with the minimum harm required for adequate protection. The resulting account

[59] Pereboom (2014), p. 168.

> would not justify the sort of criminal punishment whose legitimacy is most dubious, such as death or confinement in the most common kinds of prisons in our society. More than this, it demands a certain level of care and attention to the well-being of criminals which would change much of current policy.[60]

This justification would not undermine all practices of holding people responsible. Rather, it would call for major revisions of our current practices. We would not punish people based what they deserve, but rather incapacitate them based on the future danger they present to society. We would not make the harm any more severe than is necessary, because the criminals do not deserve to be harmed. We would need to rid our crime control system of its retributivist roots and implement a system which aims primarily to protect society.

E. Problems with the Pereboom-Style Approach

There are several significant issues with Pereboom's argument. To begin with, there is broader issue regarding the instability of the quarantine analogy. The analogy as a whole is problematic, and my discussion will shed light on why we should not justify the incapacitation of criminals using the same justification as quarantining diseased persons. But there are also several issues that would make implementing the incapacitation approach practically troublesome. First, the incapacitation theory turns a blind eye to the circumstances that actually create criminality. Second, the incapacitation theory would be largely ineffective as an actual practice. In order to illustrate my objections, I will begin by targeting the analogy to quarantine broadly and then work my way towards investigating the more particular consequences of implementation.

[60] Pereboom (2014), pp. 173-74.

Using the analogy of quarantining carriers of severe communicable diseases to bolster justifying the incapacitation of dangerous criminals is problematic. This is because quarantining is a short-term solution and Pereboom is using it as an analogy to solve a long-term problem. What I mean by referring to quarantining as a short-term solution is that quarantining is something we do in the beginning stages of a disaster before we start to implement more effective means of stopping the disease. For example, imagine that a dangerous communicable disease is spreading throughout a developed society. Initially, we might be justified in quarantining those people affected by the disease. However, continuing to quarantine those who are dangerous certainly is not viable as a long-term solution to this problem. We could not simply pluck each person out of society, one-by-one, to protect those who are not yet affected. After some initial steps like quarantining, we would need to start being proactive rather than just reactive. We would need to implement preventative measures such as administering immunizations and educating the public.

This same idea should apply to our system of crime control, as well. It is important not to just remove people from society as a reactive measure, but also to implement and emphasize proactive approaches to crime control. But what does it take to prevent people from committing crimes? Or, as Vargas might ask, what does it take to make people better moral agents? It seems as though one important factor in molding behavior is having clear, consistent rules with consistent and fair consequences for breaking those rules. However, Pereboom's incapacitation approach does not have the justificatory power to create, post, and maintain rules with pre-established consequences. This is a serious problem for crime prevention. If the rules are unclear in the first place,

then it would be very difficult to keep people from breaking those rules. Likewise, if people are unsure of the consequences that will result from breaking the rules, then those rules will be more difficult to enforce. In a report published by The Sentencing Project from November 2010, Valerie Wright explains that "in order for sanctions to deter, potential offenders must be aware of sanction risks and consequences before they commit an offense."[61] Even if hard incompatibilism is true, agents might still have the capacity to respond to reasons. But if there are widely varying punishments (ranging from nothing at all to long-term incapacitation) for the same crime based on the perceived level of future dangerousness, then the state does not offer a clear incentive to refrain from breaking the rules. This might encourage more one-time offenses where people believe they will be able to "get away with it," or small crimes that do not pose any real dangers to anyone. In many ways, the incapacitation approach is a sort of grab bag of consequences—you never know what you're going to get.

According to Pereboom's incapacitation approach, we could not justifiably take the steps required to prevent future crime. But the goal of our criminal justice system should not just be to remove dangerous people from society; rather, it should be to prevent people from committing crimes in the first place. One consequence of the incapacitation theory is that the number of people quarantined at any given time bears no reflection on how the crime control system is working as a whole. For example, a perfectly-working incapacitation system might involve 90% of the population being incapacitated to protect the remaining 10% of society, and Pereboom provides no reason to think this would be problematic. However, this would clearly be a repugnant outcome.

[61] Wright (2010), p. 5.

Rather than merely dealing with the aftermath of a disaster, we need to work on preventing the disaster from hitting in the first place.

Finally, I take issue with the way in which Pereboom reaches his conclusion. Pereboom rejects retributivism and moral education as general theories, which seems fair to me. But regarding deterrence theories, Pereboom chooses to attack one particular version of consequence-motivated justification for punishment—namely, Jeremy Bentham's utilitarianism. Because Bentham's utilitarianism entails repulsive conclusions, Pereboom rejects deterrence-inspired theories altogether. However, I argue that there is a better deterrence-inspired, consequentialist theory available. It is a theory that employs the concept of *rule* consequentialism rather than *act* consequentialism. This new rule consequentialist theory, which I will explain further in Chapter III, has the ability to hold up against hard incompatibilist threats to desert-based moral responsibility, but also avoids the moral problems Pereboom used to reject Bentham's model.

F. Conclusion

Our current system of crime control employs many retributivist-motivated beliefs and practices. If hard incompatibilism is right, then that means we cannot be held morally responsible for our actions in the basic desert sense. If we reject the concept of desert, the entire theory of retributivism is undermined. Since crime control will inevitably entail some sort of harm, we have normative obligations to provide proper justification for that harm. If we cannot justify the harm incurred from punishment on retributivist grounds, then we must either revise our conception of moral responsibility so that our practices

will be justified, or we need to change our practices to account for this lack of moral responsibility. Vargas proposes we do the former, Pereboom proposes we do the latter.

One of the primary goals of Vargas's approach was to establish a justified account of fostering better moral agency. This is something that Pereboom's view lacks. However, Pereboom takes the practice-revisionism approach, which I believe is a necessary step towards developing a justification for a system of crime control. The theory I will detail in Chapter III is able to simultaneously maintain important benefits of both Vargas's view and Pereboom's view, while avoiding the problems I noted for each. As I have argued in §II.C, our current crime control system is in dire need of substantive changes. This, I have argued, trumps the principle of philosophical conservation. However, Pereboom's argument for implementing an incapacitation theory of punishment has significant issues as well. As I explained in §II.E, the incapacitation approach would not be an effective long-term solution to controlling crime. What we need is a theory of punishment that keeps the important benefits of both Vargas's and Pereboom's views, but that avoids the problems both of those views run into. I will propose a theory which can accomplish this goal. This rule consequentialist theory of punishment emphasizes Vargas's goal of justifying practices based on their effects on creatures like us, but is created through a framework of revising our practices rather than our conceptions of moral responsibility. The following chapter will be devoted to describing and defending such a theory.

III. The Rule Consequentialist Model of Crime Control

A. Introduction

If hard incompatibilism is right, then there is a dire need to change many of our current crime control practices. Since crime control practices involve the intentional infliction of harm, such practices are morally impermissible unless they meet a high epistemic standard of justification. This means that if there is a significant enough chance the reasoning for the inflicted harm is unsound, then we ought not to inflict that harm. Hard incompatibilism entails that we cannot hold people morally responsible for their actions in the basic desert sense.[62] If we cannot hold people morally responsible in the basic desert sense, then our retributivist practices will fail to meet that high epistemic standard of justification. I explained in Chapter II that since many of our current practices are retributivist in nature, hard incompatibilism entails that those practices are unjustified. In order to remedy this problem, we need to make significant revisions to our practices.

There is an initial question of the sorts of changes that are actually available to us. If we are constructing a new criminal justice system, there are certain specifications we can change, and those we cannot. Actual world circumstances such as land area, resource availability, and human nature are more or less fixed as they are. No candidate criminal justice system should take for granted any substantial revisions to these sorts of factors because it would just not be pragmatically plausible to do so. These limitations go far beyond the realm of what is morally permissible and what is not. However, factors such as types of regulations, procedures, and repercussions for breaking the law are generally fair game. Although these factors are restricted in some ways by pragmatic limitations

[62] See §I.B. for a discussion on basic desert.

(resource availability, etc.), these are the sorts of factors which require revisions if our system of crime control is to be justified.

There have seen several different proposals for the best non-retributivist justification for crime control practices. As we saw in Chapter II, there are major flaws with all of the usual top contenders. The moral education model will not work because in most cases, it is not clear that criminals learn anything new when they are told their actions are wrong. Criminals typically already know the codes of conduct for the society in which they live, and their psychologies are generally much less malleable than those of children—for which the moral education model might actually work. But Bentham-style deterrence models do not seem to work either. It seems to have repugnant practical effects such as overly severe punishments and scapegoating. Additionally, the self-defense model of crime control is also unattractive. This model seems to only justify preventative measures and empty threats, which would not be sufficient if we want an effective system of crime control. Finally, Pereboom's view is too nearsighted. His incapacitation approach is unable to sufficiently prevent future crime from occurring because it cannot justify standardized and ensured repercussions for breaking the law. The question, then, is what to do next. If hard incompatibilism is true, then the best system of crime control will be one that (1) reaches a high epistemic standard of justification, (2) is most effective in accomplishing its goals, and (3) has the best prospects for real-world implementation and stability.

We have already seen from Vargas's approach that moral responsibility revisionism is just not enough. Since our system of crime control is already so dysfunctional and ineffective, there is not good reason to limit ourselves by the principle

of philosophical conservation. On the contrary, we actually have very good reason to make major revisions to our current system in ways that will be better for society in general. Hard incompatibilism undermines retributivism, which means much of the harm inflicted through crime control practices is unjustified. Rather than just revise our views about moral responsibility, we need to revise our practices. Following Pereboom's lead, we need to find the best crime control practices for society that also produce the least amount of harm. However, there is also an important lesson we should take away from Vargas's view: in order to achieve the best consequences for society in general, we need to implement practices which aim to make people better moral agents in the first place. The account I will lay out in this chapter accomplishes both of these goals.

My account, which I will call the *rule-consequentialism model (RCM),* appeals to rule consequentialism as a way to justify certain crime control practices. This model, I believe, accomplishes everything we want to accomplish with a good system of crime control, inflicts minimal harm, and maintains our intuitive notions of fairness and justice. I will begin my argument by giving a brief overview of consequentialism and the different forms it often takes. I will then discuss how rule consequentialism can be used as a justification for a crime control system, and explain why the RCM is successful in reaching a high epistemic standard for justifying harm. I will then discuss the implications of the RCM and what it would need to look like in application. Finally, I will consider various objections to my view and explain why each one is unfounded.

B. Rule Consequentialism as a Justificatory Theory

Before giving a detailed account of the RCM, it is important to first make a distinction between *rule consequentialism* and *act consequentialism*. According to act consequentialism, the action that agent *A* ought to do in situation *S* is the one which produces the best overall consequences. Act consequentialism, then, evaluates individual actions solely based on the consequences that will come from those actions. The good actions are the ones which produce good consequences, and the bad actions are the ones that produce bad consequences. Consequentialists disagree about what the relevant goods are to be evaluated—that is, what makes something a "good" consequence and something else a "bad" consequence. Bentham's style of utilitarianism, which we saw briefly in Chapter II, is a version of act consequentialism.[63] According to utilitarianism, the relevant good to be evaluated is general welfare. Early versions of utilitarianism viewed welfare exclusively in terms of pleasure and pain, but most modern versions of utilitarianism view welfare in broader terms.[64] Non-utilitarian consequentialist theories are also motivated by promoting welfare, but they typically believe that general welfare is not the *only* good that is relevant for evaluating actions. Other goods that non-utilitarian

[63] See Bentham (1781). He explains, "sum up all the values of all the pleasures on the one side, and those of all the pains on the other. The balance, if it be on the side of pleasure, will give the good tendency of the act upon the whole, with respect to the interests of that individual person; if on the side of pain, the bad tendency of it upon the whole. Take an account of the number of persons whose interests appear to be concerned; and repeat the above process with respect to each. Sum up the numbers expressive of the degrees of good tendency, which the act has, with respect to each individual, in regard to whom the tendency of it is good upon the whole: do this again with respect to each individual, in regard to whom the tendency of it is good upon the whole: do this again with respect to each individual, in regard to whom the tendency of it is bad upon the whole. Take the balance which if on the side of pleasure, will give the general good tendency of the act, with respect to the total number or community of individuals concerned; if on the side of pain, the general evil tendency, with respect to the same community" (pp. 32-33).

[64] This development occurred because there seem to be things in the world we care very deeply about that are not directly related to pain or pleasure. We often care deeply, for example, about the wellbeing and success of our friends and family—even if it has no effects on our own personal level of pleasure.

consequentialists tend to favor include fairness and justice.[65] However, this is not to exclude welfare as a relevant good; these are goods in *addition* to welfare that are relevant for evaluating consequences.

Rule consequentialism is different from act consequentialism in important ways. Rather than assessing individual actions based on their consequences, rule consequentialism promotes particular *rules* which are justified by their consequences. Whereas act consequentialism evaluates particular acts, rule consequentialism evaluates classes of acts. This theory assesses individual actions based on whether they follow these rules. For example, let's assume that general acceptance of the rule *feed your children healthful meals* would have positive overall consequences: children would generally be healthier, they would grow into happier people, and there would be fewer cases of malnutrition-related illnesses. The general acceptance of this rule would promote the overall welfare of society. The rule would be best formalized in a way that is fair and just. Thus, rule consequentialism would determine that the rule *feed your children healthful meals* is a good rule. If agent *A* were to break that rule by only feeding his children candy and soda every day, then *A* would be doing something morally wrong. According to act consequentialism, however, *A*'s action is not wrong in virtue of the act's consequences. Rather, *A*'s action is wrong because it breaks the rule *feed your children healthful meals*, which is justified based on its consequences.

Consequentialism is a metaethical theory. It is typically used to decide what's right or wrong, what ought or ought not be done. It establishes a moral code. However, I want to remain neutral about consequentialism as a metaethical theory. The way in which I will use rule consequentialism does not require that we accept any sort of

[65] See Broome (1991), Feldman (1997), Kagan (1999), and Hooker (2005).

consequentialism in our metaethics. Instead, I plan to use rule consequentialism as a framework for a justified system of crime control. It does the justificatory work for intentionally inflicting harm on criminals. Like retributivism, or the moral education model, or Bentham-style utilitarianism, or Pereboom's incapacitation approach, the RCM aims to provide proper justification for a particular system of crime control. In Chapter II, I explained that in order to maintain a system of crime control that regularly inflicts harm, we need to accomplish two goals. First, we must find proper justification for the harm inflicted to control crime; second, we must determine the extent to which such harm is justified. For the rest of this section, I will focus on the first of these goals. I will address the second question in §III.C.

Any pragmatically useful system of crime control will involve intentionally inflicting some amount of harm. This is because preventing crime requires restricting certain freedoms. However, in a society like ours, it is better to have an effective system of crime control and criminal law procedures in place than to not. Still, whatever harm we allow needs to be justified. The RCM provides a framework for this justification. The framework is this:

> **RCM:** A system of crime control consists of laws (regulations, procedures, and repercussions). The harm produced by the Rule Consequentialist Model of crime control is justified because those laws aim to produce the best overall consequences. The best overall consequences are the consequences which maximize general welfare and promote fairness and justice. If these good consequences outweigh the harm produced by the laws as much as reasonably possible, then that harm is justified.

The RCM justifies the harm produced by crime control because it implements rules based on the expected consequences of implementing and enforcing those rules. The goods relevant to the expected consequences include general welfare, fairness, and justice. This

means that the laws enacted will aim to maximize welfare and promote fairness and justice. One important thing to notice is that these are typically the goods we aim to secure through our current crime control system. The main difference between the goals of the RCM and the goals of our current system is that retributivism does not play any part in the RCM.

There are two points I should make about fairness and justice. My first point is that fairness and justice are not undermined by hard incompatibilism. That is, we can still rationally value fairness and justice even if we cannot hold people morally responsible in the basic desert sense. After all, the reason we cannot inflict harm on criminals without proper justification in the first place is because it would be unjust and unfair to do so. Principles of fairness encourage policies such as governmental transparency, procedural due process, and predictable repercussions. Fairness, then, demands consistency and stability within the law. Principles of justice encourage policies such as the rule of law, substantive due process, and legal accountability. Justice, then, demands a degree of impartiality and equality. We do not need basic desert for these ideas to be rational.

Even without basic desert, maintaining these values is desirable both from the standpoint of the citizens and the lawmakers. This brings me to my second point: it's not just the case that we *can* maintain these values if hard incompatibilism is true, but it's also the case that we *should* maintain these values. The policies listed above may not always maximize welfare. Nonetheless, fairness and justice are important components for a system of crime control if that system is going to be attractive to its citizens, sustainable, and effective. We need to balance out welfare maximization with these ideals of fairness and justice because that is what makes a system of criminal law a good system of

criminal law. These are the values that we want to gain from having a system of criminal law in place. Implementing regulations, procedures, and repercussions that are fair and just gives ordinary citizens a sense of assurance. Citizens know what to expect from the system, and they know what's expected of them.

I, like Vargas, believe that our practices of holding one another responsible should derive justification from their effects on creatures like us. In particular, I believe that our system of criminal law can only be justified if it aims to produce the best overall consequences. The best way to make this happen through a system of criminal law is to have effective rules in place that aim to maximize welfare while balancing principles of fairness and justice. Essentially, we want rules that will do the best job of allowing individuals to live their lives as they please without unreasonably disrupting the lives of others, and we want people to follow those rules. Because of this, the types of rules that would be required by the RCM would be ones that aim towards making people better moral agents—or at the very least, agents who follow the rules of society. Although crime control systems will inevitably involve some amount of intentional, state-sanctioned harm, the RCM provides a good framework for justifying that harm. Harm is justified because the best consequences come about by inflicting that harm. We now must ask how *much* harm is justified through the RCM, and what the RCM would look like in application. The next section addresses these questions.

C. Applying the Rule Consequentialist Model

When considering the sorts of rules that would be justified through the RCM, it is important to keep in mind the overarching goal: to maximize welfare while promoting

fairness and justice. This goal should be the foundation of all regulations, procedures, and repercussions within the system. To do this, there are a few general tactics that the RCM prescribes. First, the RCM would support inflicting as little harm as possible while maintaining its effectiveness. Repercussions for breaking the laws should be no harsher than is necessary to obtain the optimal consequences. If there are two potential repercussions for breaking a rule, and both of these rules have the same deterrent effect, then according to the RCM we are morally required to choose the rule that does the best job of maximizing welfare. This means that, ceteris paribus, we are morally required to select the one which would inflict the least amount of harm on the criminal. After all, the RCM ultimately aims to mold better moral agents—not punish those who have strayed. Because of this, the RCM would also support the use of rehabilitation techniques. The goal of rehabilitation is to make the criminal a better moral agent and prevent any future harm, so this sort of technique should be used whenever possible. But the welfare of the criminals is obviously not the only welfare that matters. Maximizing the welfare of non-criminals is also very important for the RCM.

The second tactic that the RCM would support is utilizing effective deterrence methods of crime control. Deterrence means keeping people safe, stopping people from infringing on each other's rights, and maintaining an environment which allows people to flourish and achieve their own personal goals (to a reasonable extent). Simply put, less crime leads to greater general welfare. Effectively deterring crime is key for the RCM's success. Luckily, these two tactics—maximizing the welfare of criminals and maximizing the welfare of non-criminal society—are not contradictory. In fact, these two goals are much more compatible than they might first appear to be.

There has been a significant amount of research over the past few decades that indicates certainty of punishment has a much greater deterrent effect than severity of punishment does.[66] What's more, increasing the severity of punishment actually has adverse effects on the certainty of punishment. As Mark Kleiman explains in his book on this subject, "severity, at least in the form of lengthy prison and jail stays, is the enemy of certainty and swiftness."[67] Kleiman makes this connection because of several factors. First, there is the simple factor of available facility space. If we pushed (as we have been in the US) for longer and harsher sentences, then there would simply not be enough resources to incarcerate everyone who ought to be incarcerated. Prisons around the US are overpopulated and overcrowded as it is. As stated above, land area and public resources are not the sorts of things we can actually make substantial changes to for the sake of a justified system of crime control. This point, then, boils down to simple math. One twenty-year sentence of incarceration uses up the same amount of space over time as forty consecutive six-month sentences. If there were shorter sentences, we would be better equipped to dole out more certain consequences. But in addition to insufficient space, Kleiman offers another reason why an increase in severity often means a decrease in certainty: "more severe punishments will be more fiercely resisted."[68] This is a mere observation about human nature. The harsher the repercussions are, the less willing criminals will be to just accept the repercussions.

Less severe punishment maximizes welfare in two ways. As we have just seen, less severe repercussions allow for more certain repercussions, which creates a deterrence

---

[66] Wright (2010), von Hirsch *et al.* (1999), Nagin and Pogarsky (2001), Kleiman (2009) especially ch. 6.
[67] Kleiman (2009), p. 93.
[68] *Ibid.*

effect. With greater deterrence, there is less crime. This decrease in crime maximizes the welfare of those who would have been victims of crimes, but it also promotes a greater sense of safety and trustworthiness throughout society. In addition to a decrease in crime, administering less severe punishment also allows us to induce a smaller amount of harm on people that do commit crimes. This in itself helps to increase welfare, because there is less harm in the world. But less severe punishment is also better for criminals down the line, too. In a meta-analysis published in 1999, after controlling for other relevant factors, longer prison sentences were actually shown to *increase* the chances of recidivism. Bruce Western explains how long-term imprisonment can impact a criminal even after release:

> Imprisonment is an illegitimate timeout that confers an enduring stigma. Employers of less-skilled workers are reluctant to hire men with criminal records. The stigma of a prison record also creates legal barriers to skilled and licensed occupations, rights to welfare benefits, and voting rights. Later chapters will show that ex-prisoners earn lower wages and suffer more unemployment than similar men who have not been incarcerated. Former prisoners are also less likely to get married or live with the mothers of their children. By eroding opportunities for employment and marriage, incarceration may also lead inmates back to a life of crime.[69]

This shows us that less severe punishment allows us to not only increase the certainty of repercussions, but it also helps to prevent future crime. Now, I am not arguing that we should merely give criminals a slap on the wrist for egregious crimes. Clearly, if the repercussions are too insignificant, then the deterrence effect will diminish. What I am arguing is that the severity of the repercussions should not be *any more* severe than is necessary to produce the best overall consequences. This likely means that sometimes the best repercussions to put in place for breaking a rule will be a type of rehabilitation or community service. But whatever the repercussions, they should be as certain as possible and no more severe than necessary.

---

[69] Western (2006), p. 21.

The third tactic the RCM would employ is promoting maximal compliance with the rules. Before we can even get to utilizing deterrence methods, people need to know and understand what the rules of society are. This involves making the rules as easy to learn and understand as possible, and also making the rules easy to follow. To do this, the rules for the RCM would be widely available and readily accessible. The rules would also favor simplicity. The less complicated the system of rules are, the easier it will be for people to follow the rules. This tactic ties together nicely with deterrence. If the rules are clear and relatively simple, and the repercussions for breaking the rules are certain, then we have ideal conditions for maximal compliance.

The RCM gives us a framework for justifying the harm induced by crime control practices. Each tactic that the RCM prescribes is aimed at maximizing welfare. But in order for this system to be sustainable, we would also need to do our best to promote fairness and justice through the rules we create. A large part of producing the best consequences involves guiding people's behavior, which maintains Vargas's focus of making people better moral agents. But the RCM takes this goal several steps further by creating a system which specifically aims to produce the best overall consequences. Our system of crime control needs more than just a revision in our theories; it needs tangible change. And unlike Pereboom's view, the RCM is a long-term solution. Pereboom's view does little to stop crime from occurring in the first place, and does not permit the exact features which has been shown to reduce crime: rules that are definitive and repercussions that are certain. However, there are still several objections to the RCM that will be important to consider. The next section turns to a discussion of these objections.

D. Objections and Responses

The first objection I will consider is a version of a common objection to rule consequentialism as a metaethical theory. It is also, I believe, one of the strongest objections against rule consequentialism. It goes something like this: if the rules are created to produce the best consequences, then why should we follow the rules in a case where breaking the rules would produce the best consequences? Now, it is important to distinguish exactly what this objection is arguing—and more particularly, what it is not arguing. What this objection is *not* arguing is that there is an internal inconsistency with rule consequentialism. It is not a contradiction when this phenomenon occurs. In fact, rule consequentialism intentionally diverges with act consequentialism in many cases; this is an important aspect of the view. This objection, then, is *not* saying there is something inherently flawed with individual cases in which following the rule would not produce the best consequences. Instead, this objection states that it is counterintuitive or unattractive to have instances where the morally right thing to do according to rule consequentialism would be the thing that produces sub-optimal consequences, when the whole metaethical system aims to produce the best consequences. In other words, if the goal is to produce the best consequences, then why wouldn't we just choose our actions based on whatever would produce the best consequences? This is where it is important to remember that I am proposing a legal theory rather than a moral theory. The RCM does not determine which actions are right and which are wrong. Rather, it sets up a framework for controlling crime that is justified and aims to produce the best overall consequences.

We as humans have very limited epistemic capacities. When it comes to cognizing, there is a lot we cannot do. We cannot, for example, consider every possible option available to us at any given time. We also cannot accurately predict the exhaustive list of consequences of our actions. Because of such limitations, we get things wrong a significant amount of the time. If the RCM had some sort of caveat like *"...unless breaking this rule would produce better consequences than following it,"* then people would need to be constantly deciding whether or not to break the rule. Thus, rule consequentialism would collapse into act consequentialism. Not only would this mean that all of the typical objections to act consequentialism would apply,[70] but it also means that the actual rule would no longer be effective. If the rule is not effective, then it cannot be expected to produce the best consequences. People could constantly be making the incorrect decisions about which actions would produce the best consequences. Rules which are created through the RCM in particular aim to produce the very best consequences out of other possible rules. So, rather than trying and sometimes failing to decide which action would produce the best consequences, rules give us a default option that aims to produce the best consequences overall. As long as people follow the rules, then overall, welfare will be maximized and the principles of fairness and justice will be maintained.

The RCM keeps us from getting things wrong when we are making decisions about our actions. But what about cases in which it is clear that breaking the rule would certainly produce better consequences than following the rule? Shouldn't the RCM allow us to break the rules in those particular cases if our goal is to produce the best consequences? The answer is no, simply because having a caveat like the one above

[70] See §II.D.

would be disastrous for a legal system. Along with maximizing welfare, the RCM must balance principles of fairness and justice as well. Think about it in terms of our current legal system—just because you have good reasons to break the law does not mean it is permissible to break the law. Even if Robin Hood has good reason to steal from the rich, our legal system should not make exceptions for Robin Hood-type cases of stealing. Rather than having clear, simple, easy-to-follow rules, these sorts of exceptions would create chaos and confusion. Applying the rules would not be fair and consistent, it would be convoluted and unstable. The question is not whether particular actions would be morally permissible or not. Rather, the question is what the effects would be if we allowed some people to break the law under some circumstances. I have argued that the effects would be overall bad for society. This is, after all, part of the reason why I am suggesting a rule consequentialist theory rather than an act consequentialist theory. The rules of a legal system need to be clear, consistent, impartial, equally applied, and stable if they are going to produce desirable results. We can do this with a rule consequentialist system; we cannot do this with an act consequentialist system. It is a much better systemic practice to have clear rules and clear repercussions for breaking the rules.[71]

Another possible objection to the RCM involves crimes that are not typically able to be deterred. For example, spousal murders and other crimes of passion have often been considered to be undeterrable.[72] This is because they are typically not pre-meditated acts, which means that aggressors are not able to engage in the sort of rational deliberation that

[71] Of course, as with our current system, there is no reason why the RCM wouldn't allow us to consider mitigating factors when assessing the criminal's case. It's not as if we would need the exact same consequences to be inflicted for breaking a particular rule in any circumstance. Even in a hard incompatibilist rule, people are able to respond to reasons. The RCM is able to account for people having good reason to do something. This will not necessarily get them off the hook entirely, but mitigating factors could change the consequences of breaking rules under certain circumstances.

[72] Shepherd (2004).

deterrence appeals to.[73] The RCM is essentially a forward-looking system of crime control. This means that its focus is on producing certain effects for the future rather than reacting to what has been done in the past. Retributivism would be an example of a backwards-looking system of crime control. But since the RCM is forward-looking in this way, rules should only be implemented if their implementation will aim to produce the best overall consequences. A rule that says *do not commit spousal murders* would not be effective because (1) if the rule was not in place, it is unlikely that many people would murder their spouses *just because* there is no such rule in place and (2) if the rule were in place, it would not deter people from committing the crime at all. The people who were not going to kill their spouses before the rule was implemented will continue to not kill their spouses after the rule was implemented, and the people who do kill their spouses would have done so regardless of whether there were a rule against it. This means that the rule's implementation is not actually effective. Implementing this rule (and repercussions for breaking that rule) would then only add harm to the world: no additional good comes from implementing the rule, but some harm occurs when offenders are forced to face repercussions. Because of this, the RCM would not be able to implement laws such as *do not commit spousal murders.* The objection then, goes something like this: any system of crime control that allows spousal murders is not a very good system. At the very least, it becomes a much less attractive model when you start thinking of all the similar undeterrable crimes that the RCM would be required to allow.

Let's assume that it is correct to say that some crimes are undeterrable. Even so, I do not believe the RCM would require that spousal murders, crimes of passion, or other similarly undeterrable crimes be permissible. Although the RCM does aim to maximize

[73] Glaser (1977).

welfare, maximizing welfare must be balanced against principles of fairness. Since fairness demands consistency and stability within the law, the RCM would discourage making these sorts of distinctions between types of murder. If the act of murder is considered wrong, and it generally has bad effects on society, then our system of crime control should not allow it. Picking out all the individual cases and types of murder, and breaking them into their own individual rules would make the system unnecessarily complex. But the RCM encourages simplicity in its rules. If we want to prohibit the act of murder, then we should prohibit the act of murder. Reasons for breaking those rules should still be considered as potential mitigating factors in deciding repercussions, but this does not mean that murder should be explicitly permissible in particular sorts of cases. Since the RCM would be implemented and enforced by people, there are also inevitably many limitations that come with evaluating a person's reasons for acting. There is not an easily applicable way to distinguish between *doing X because of Y* and *doing X because of Z.* Rather than making these distinctions, we should just prohibit *doing X*. This is much easier to enforce, and it makes the rules easier to understand and follow. The RCM aims to produce rules that create the best overall consequences, and it does not seem at all clear to me that explicitly allowing acts such as spousal murders would produce the best consequences.

A third objection one might have to the RCM is that we would never be able to create a set of rules that actually produce the best consequences. After all, since we are such limited beings, we cannot consider the consequences of every possible rule. Furthermore, we cannot see into the future to determine what the actual consequences of the rules will be. Since the rules of this system will have been created by humans, it is

doubtful that all rules (or perhaps even any rules) will attain the perfect balance between welfare-maximization, fairness, and justice. However, I am not suggesting that the RCM would be perfect. Any system that is run by non-perfect beings in a non-perfect world is bound to be a non-perfect system. And I certainly do not expect that implementing the RCM would get it right the first time. On the contrary, I expect that the RCM would need continual revisions. We should be constantly trying to make our system better. When there is new information, we should adjust our system accordingly. When there are better procedures discovered, we should change the procedure that we currently have in place. The RCM's *aim* is to produce the best overall consequences, but it is up to the people creating and enforcing the laws to figure out exactly how we achieve that.

E. Conclusion

If we want some way to control crime in our society, we need to cause some amount of harm. The Rule Consequentialism Model gives us a framework to justify that harm. According to the RCM, we should implement rules (including the consequences for breaking those rules) based on the consequences they are expected to bring about. In deciding which rules to implement, we should try to maximize welfare while balancing fairness and justice in society. The RCM would deter crime by implementing swift and certain repercussions for breaking the rules. It would also deter crime by making the rules clear and publicly available. This involves making rules that are as easy to learn and understand as possible, and also relatively easy to follow. The RCM would also forbid treating criminals any more harshly than is necessary to produce the optimal consequences. Because of this, and the ultimate forward-looking goal of maximizing welfare, the RCM would also favor practices of rehabilitation whenever possible.

The RCM maintains Vargas's goal of building better moral agents. We want to produce the best consequences, and this can only happen if people are not constantly infringing on one another's rights; this can only happen if people are good moral agents. However, the RCM goes several steps past Vargas's approach. Because our current crime control system is largely unjustified due to retributivism being undermined by incompatibilism, we have significant changes to make to our current system. However, Pereboom's incapacitation approach just won't accomplish our goals. This approach fails to make any progress towards making people better moral agents, and it does not allow for implementing the type of consequences that actually work: determined consequences that are certain to happen. Pereboom's view is simply not a long-term solution. Thus, the RCM is the best system of crime control because it (1) reaches a high epistemic standard of justification, (2) is most effective in accomplishing its goals, and (3) has the best prospects for real-world implementation and stability.

Conclusion

Whether the hard incompatibilists are right is not a settled matter. In Chapter I, I laid out the problem of free will and moral responsibility as Pereboom describes the issues. I explained why it is unlikely that we can hold people morally responsible in the basic desert sense if causal determinism is false, and why it is unlikely that we can hold people morally responsible in the basic desert sense if causal determinism is true. I have supported Pereboom's conclusion that compatibilism is not a viable option. However, in a 2009 survey organized by David Bourget and David Chalmers, 69.4% of philosophers surveyed identified as compatibilists about free will.[74] That is a very significant portion of the philosophical population. But I believe that Pereboom's four-case manipulation argument shifts the burden of proof to the compatibilists. The compatibilists need to come up with a good response that shows why Pereboom's four-case manipulation argument is unsound. As things are, I believe we have good reason to believe the hard incompatibilists are right.

Since we have good reason to believe the hard incompatibilists are right, then there is good reason to reevaluate our responsibility-characteristic practices. The practices that I was concerned with in this piece were those included in our system of criminal law. If we cannot hold people morally responsible in the basic desert sense, then our retributivist practices will fail to meet that high epistemic standard of justification. Since many of our current practices are retributivist in nature, hard incompatibilism entails that those practices are unjustified. In order to remedy this problem, we need to make significant revisions to our practices.

---

[74] Bourget and Chalmers (2014)

If hard incompatibilism is true, then the best system of crime control will be one that meets the standard of a three-pronged test. First, the harm produced by that system must reach a high epistemic standard of justification. The RCM justifies the harm produced by crime control because it implements rules based on the expected consequences of implementing those rules. This reasoning, I believe, reaches a high epistemic standard of justification. Second, the system must be most effective in accomplishing its goals. The RCM is structured in a way that effectiveness is one of its primary goals. Laws will be specifically designed to deter crime and promote general compliance with the rules. Third, the system must have the best prospects for real-world implementation and stability. I believe that the RCM has better prospects for real-world implementation than any of the other models we've seen. Out of the most promising options to replace retributivism—moral education, Bentham-style deterrence, self-defense, and Pereboom's incapacitation approach—the RCM is the best option we have. The RCM provides a framework for justifying a particular system of criminal law. This model, I believe, accomplishes everything we want to accomplish with a good system of crime control, inflicts minimal (and justifiable) harm, and maintains our intuitive notions of fairness and justice.

If we cannot be held morally responsible for our actions, then I believe we should move towards a system like the RCM. Since libertarianism is either implausible or irreconcilable with moral responsibility, it is up to the compatibilists to show how we can salvage the sort of free will required for moral responsibility in the basic desert sense. But if they cannot do this in a way that satisfactorily overcomes the four-case manipulation argument, then we have good reason to implement the RCM.

References

Antunes, G. & Hunt, A. L. (1973). "The Impact of Certainty and Severity of Punishment on Levels of Crime in American States: An Extended Analysis." *Journal of Criminal Law and Criminology*, *64*(4), 486-493.

Ashworth, A. (2000) "Is Criminal Law a Lost Cause?" 116 *Law Quarterly Review*, *116*(2), 225-256.

Balaguer, M. (2004). "A coherent, naturalistic, and plausible formulation of libertarian free will." *Noûs*, *38*(3), 379-406.

Bentham, J. (1781 [2000]). *An Introduction to the Principles of Morals and Legislation*. Batoche Books.

Bentham, J. (1843). *Principles of Penal Law.* W. Tate.

Berofsky, B. (2002). "Ifs, cans, and free will: The issues." In Robert Kane (ed.), *The Oxford Handbook of Free Will,* New York: Oxford University Press, 181-201.

Blumstein, A. (1982). "On the racial disproportionality of United States' prison populations." *J. Crim. l. & Criminology*, *73*, 1259.

Bourget, D. & Chalmers, D. J. (2014). "What Do Philosophers Believe?" *Philosophical Studies* 170: 465-500.

Broome, J. (1991). "Utility." *Economics and Philosophy*, *7*(1):1-12.

Campbell, J. (1997). "A compatibilist theory of alternative possibilities." *Philosophical Studies*, *88*(3), 319-330.

Chisholm, R. (1966) “Freedom and Action.” In K. Lehrer (ed.) *Freedom and Determinism*, New York: Random House, 11-44.

Clarke, R. (1993). "Toward a credible agent-causal account of free will." *Noûs*, 191-203.

Dolinko, D. (1997). "Retributivism, consequentialism, and the intrinsic goodness of punishment." *Law and Philosophy*, *16*(5), 507-528.

Durose, M., Cooper, A., & Snyder, H. (2014). *Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010*, Bureau of Justice Statistics Special Report, April 2014, NCJ 244205.

Ekstrom, L. W. (2000). *Free will*. HarperCollins Publishers.

Farrell, D. M. (1989). "Intention, Reason, and Action." *American Philosophical Quarterly*, 283-295.

Feldman, F. (1997). *Utilitarianism, hedonism, and desert: Essays in moral philosophy*. Cambridge: Cambridge University Press.

Fischer, J. M. (1994). *The Metaphysics of Free Will: An Essay on Control*. Cambridge: Blackwell.

Fischer, J. M. (2002). "Frankfurt-type examples and semi-compatibilism." In Robert Kane (ed.), *The Oxford Handbook of Free Will*. Oxford University Press.

Frankfurt, H. (1969). "Alternate possibilities and moral responsibility." *Journal of Philosophy*, 829-839.

Frankfurt, H. (1971). "Freedom of the Will and the Concept of a Person." *Journal of Philosophy, 68*(1), 5-20.

Freudenberg, N. (2001). "Jails, prisons, and the health of urban populations: a review of the impact of the correctional system on community health." *Journal of Urban Health, 78*(2), 214-235.

Ginet, C. (1990). *On action*. Cambridge University Press.

Ginet, C. (1996). "In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing." *Philosophical Perspectives* 10:403-17.

Ginet, C. (2007). "An Action Can be Both Uncaused and Up to the Agent." In Lumer (ed.), *Intentionality, Deliberation, and Autonomy*. Ashgate. 243-255.

Glaser, D. (1977). "The Realities of Homicide versus the Assumptions of Economists in Assessing Capital Punishment." Journal of Behavioral Economics *6*(1). 243-269.

Goetz, S. (2000). "Naturalism and libertarian agency." In W. L. Craig & J. P. Moreland (eds.), *Naturalism: A Critical Analysis*. Routledge.

Griffith, M. (2010). "Why agent-caused actions are not lucky." *American Philosophical Quarterly*, 43-56.

Haji, I. (1998). *Moral appraisability*. Oxford University Press.

Hooker, B. (2005). "Fairness." *Ethical Theory and Moral Practice* 8 (4):329 - 352.

Kagan, S. (1999). "Equality and desert." In L. P. Pojman & O. McLeod (eds.), *What Do We Deserve?: A Reader on Justice and Desert*, 298-314. Oxford University Press.

Kane, R. (1995). "Control, Responsibility and Free Will." *Southwest Philosophy Review* 11 (2):255-258.

Kane, R. (2005). *A contemporary introduction to free will*. Oxford University Press.

Kleiman, M. (2009). *When brute force fails: How to have less crime and less punishment*. Princeton University Press.

Levy, N. (2012). "Skepticism and sanction: The benefits of rejecting moral responsibility." *Law and Philosophy*, *31*(5), 477-493.

Levy, N. & McKenna, M. (2009). "Recent work on free will and moral responsibility." *Philosophy Compass*, *4*(1), 96-133.

Mauer, M. (2001). "The Causes and Consequences of Prison Growth in the United States." *Punishment & Society*, *3*(1), 9-20.

McCann, H. (1998). *The works of agency: On human action, will, and freedom*. Cornell University Press.

McKenna, M. (2001). "Source incompatibilism, ultimacy, and the transfer of non-responsibility." *American Philosophical Quarterly*, 37-51.

McKenna, M. (2008). "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument." *Philosophy and Phenomenological Research*, *77*(1), 142-159.

Mele, A. R. (1995). *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.

Mele, A. R., & Robb, D. (1998). "Rescuing Frankfurt-style cases." *Philosophical Review*, 97-112.

Nagin, D. & Pogarsky. G. (2001). "Integrating Celerity, Impulsivity, and Extralegal Sanction Threats into a Model of General Deterrence: Theory and Evidence," *Criminology*, 39(4).

Nichols, S. (2013). "Brute retributivism." *The future of punishment*, 25-46.

Pereboom, D. (1995). "Determinism al dente." *Noûs*, 21-45.

Pereboom, D. (2001). *Living without free will*. Cambridge University Press.

Pereboom, D. (2008). "A Hard-line Reply to the Multiple-Case Manipulation Argument." *Philosophy and Phenomenological Research*, *77*(1), 160-170.

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

Pound, R. (1927). "Introduction." In F. B. Sayre, *A Selection of Cases on Criminal Law*. Lawyers Co-Operative Publishing Co.

Robinson, P. H. & Darley, J. M. (2015). "Does Criminal Law Deter? A Behavioural Science Investigation." Oxford J Legal Studies (Summer) 24 (2): 173-205.

Roskies, A. (2006). "Neuroscientific challenges to free will and responsibility." *Trends in cognitive sciences*, *10*(9), 419-423.

Roskies, A. L. (2012). "How does the neuroscience of decision making bear on our understanding of moral responsibility and free will?" *Current opinion in neurobiology*, *22*(6), 1022-1026.

Rowling, J.K. (2005). *Harry Potter and the Half-Blood Prince.* Arthur A. Levine Books.

Sartorio, C. (2011). "Actuality and responsibility." *Mind*, *120*(480), 1071-1097.

Shepherd, J. M. (2004). "Murders of Passion, Execution Delays, and the Deterrence of Capital Punishment," *Journal of Legal Studies*, *33*(2): 283-323.

Strawson, G. (2000). "The unhelpfulness of determinism." *Philosophy and Phenomenological Research* 60 (1):149-56.

Strawson, G. (2004). "Free Agents." *Philosophical Topics* 32 (1/2):371-402.

Strawson, G. (2011). "The impossibility of ultimate responsibility?" In R. Swinburne (ed.), *Free Will and Modern Science*. Oxford University Press/British Academy.

Stump, E. (1996). "Libertarian freedom and the principle of alternative possibilities." *Faith, Freedom, and Rationality: Philosophy of Religion Today.* Lanham: Rowman & Amp; Littlefield. 73-88.

Stump, E. (1999). "Dust, Determinism, and Frankfurt." *Faith and Philosophy*, *16*(3), 413-422.

Tadros, V. (2011). *The ends of harm: The moral foundations of criminal law*. Oxford University Press.

Taylor, R. (1966). *Action and Purpose*. Prentice-Hall.

Tonry, M. (2013). "Sentencing in America, 1975–2025." *Crime and Justice*, *42*(1), 141-198.

van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press.

Vargas, M. (2013). *Building better beings: A theory of moral responsibility*. Oxford University Press.

von Hirsch, A. *et al.* (1999) “Criminal Deterrence and Sentence Severity: An Analysis of Recent Research,” Oxford: Hart Publishing.

Western, B. (2006). *Punishment and inequality in America*. Russell Sage Foundation.

Widerker, D. (2006). "Libertarianism and the philosophical significance of Frankfurt scenarios." *The Journal of Philosophy*, 163-187.

Wright, V. (2010). *Deterrence in criminal justice: Evaluating certainty vs. severity of punishment*. Sentencing Project.