OBJECTIONS TO OPTIMISTIC

FREE WILL SKEPTICISM

by

Hannah Vu

Submitted in partial fulfillment of the

requirements for Departmental Honors in

the Department of Philosophy of

Texas Christian University

Fort Worth, TX

December 13, 2021

OBJECTIONS TO OPTIMISTIC

FREE WILL SKEPTICISM

Project Approved:

Supervising Professor: Kelly McCormick, Ph.D.

Department of Philosophy

John Harris, P.h.D.

Department of Philosophy

Michael Strausz, P.h.D.

Department of Political Science

ABSTRACT

The free will debate is fundamentally grounded in asking whether people have the sort of control required for moral responsibility given physical facts about the universe and other causal factors involved with our actions. This paper surveys the main views in the contemporary free will debate and examines how Derk Pereboom argues that under any view, we never have the control required for moral responsibility. The final section of this paper examines the views of two prominent optimistic free will skeptics, Derk Pereboom and Gregg D. Caruso, who argue that we lack free will and that we would be better off if we accepted this fact and its implications. I will ultimately argue that features of these skeptical views have pragmatic and moral concerns that lead us to reject some of their implications.

Contents

Objections to Optimistic Free Will Skepticism

Introduction

The free will debate is fundamentally grounded in asking whether people have the sort of control required for moral responsibility, given physical facts about the universe and other causal factors involved with our actions. In light of this, many define free will as "the strongest sort of control required for moral responsibility."[1] In other words, moral responsibility and basic desert of praise and blame presuppose that we have free will. This paper focuses on examining the views of forward-looking free will skeptics who argue that we lack free will and that we would be better off if we accepted this fact and its implications. I will ultimately argue that two prominent optimistic skeptical views have pragmatic and moral concerns that lead us to reject some of their implications.

In the first section of this paper, I survey the main views in the free will debate. In Section 2, I discuss Derk Pereboom's hard incompatibilism, which holds that both determinism and indeterminism are incompatible with our ideas of free will and moral responsibility. In Section 3, I explain some of the main implications of free will skepticism, especially regarding our attitudes and theories of punishment. In Section 4, I examine quarantine theories put forth by Derk Pereboom and Gregg D. Caruso. Finally, in Section 5, I present my own objections to free will skepticism. Here I will question optimistic free will skeptics' arguments about reactive attitudes and the moral implications of quarantine-like models for criminal punishment. I believe these negative arguments persuasively motivate an illusionist view and encourage us to maintain

---

[1]Mele, A. R. (2008). *Free Will and Luck.* Oxford University Press.

our belief in free will, moral responsibility, and basic desert despite the theoretical force of skeptical arguments.

1. Surveying Views in the Contemporary Free Will Debate

The free will debate includes both metaphysical questions about whether free will exists and ethical and normative questions about moral responsibility, agency, basic desert, blame, praise, punishment, and many other concepts. One of the metaphysical issues that prominently features in the free will debate asks whether free will is compatible with determinism. Peter Van Inwagen defines determinism as a conjunction of two thesis: (1) that for every given moment, there is only one physically possible way that things can go, and (2) that facts about the past and the laws of nature entail only one distinct future.[2] Those who believe that we do have free will regardless of whether determinism is true or not are compatibilists. People who believe that determinism and free will are incompatible are incompatibilists. Among incompatibilists, there are two kinds: libertarian incompatibilists and free will skeptics. Libertarian incompatibilists believe that we have free will and that determinism is false. Free will skeptics believe that determinism and free will are incompatible, but unlike their libertarian counterparts, take seriously the possibility that we lack free will. I will offer a brief survey of some of the contours of each of these views of free will.

1.1 Compatibilist Views

Compatibilists attempt to show that someone can be morally responsible even if the universe is deterministic. In contrast, incompatibilists typically subscribe to the Principle of Alternative Possibilities, or PAP; PAP holds that for a person to have free will, then it must be

---

[2] Van Inwagen, P. (1975). The Incompatibility of Free Will and Determinism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 27*(3), 186.

the case that they could have done otherwise.[3] Most compatibilists deny PAP. That is, they do not

believe we need genuine and physically possible alternatives in order to be candidates for moral

responsibility. Many compatibilists argue that what moral responsibility requires is that our

actions are connected to us in a certain way, not that the universe is indeterministic.

Here I will discuss two of the most prominent kinds of compatibilist views in the free will

debate: hierarchical and reasons-responsive views.

Harry Frankfurt's hierarchical view is one of the most prominent compatibilist accounts

in the free will debate. Frankfurt argues directly against PAP by famously providing what we

now refer to as Frankfurt-style counterexamples.[4] Frankfurt's original example shows that even

though agents do not have physically open alternatives, we can still intuitively think they are

morally responsible. Frankfurt shows this in one case where an agent named Jones decides to

perform an action for his own reasons. Another person, Black, has manipulated Jones in such a

way that at any given moment he can control Jones's actions if he wants to, unbeknownst to

Jones. Jones 4 decides to and carries out a morally wrong action on his own, and Black never has

to use his manipulation tactics. Here is an example of a case where we can see that Jones did not

genuinely have open options: he could have taken the action on his own, or Black would have

manipulated him to perform the action. There is no physically possible way that Jones could

have avoided doing what Black asked. However, we still intuitively feel like Jones was not

necessarily coerced and that he is still morally responsible for his action (834-836).

Frankfurt believes that what is needed for moral responsibility is not physically open

alternatives and argues that we should replace PAP with the following principle: "a person is not

---

[3] Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy, 66*(23), 829-839.
[4] Ibid.

morally responsible for what [they] have done if [they] did it only because they could not have done otherwise, or *because* [they] had to do it" (838). If we accept this principle, then Jones 4 is morally responsible since he performed the action on his own because of his own reasons, not because he was forced to or because he had no other option. What matters for Frankfurt is what he calls our "second-order desires," or our desires about our desires (6). For example, take a person who wants to eat a slice of chocolate cake. However, they wish that they did not want to eat the cake, because they want to start dieting. Their first order desire is to eat cake, while their second order desire is that they wish their first order desire was different. If the person succeeds in not eating the cake, then their second order volition is effective. Frankfurt argues that it does not matter how we came to have our wills or desires. They could be determined and we can still be morally responsible, so long as we have the wills we want and the power to realize them in action.

A common objection for Frankfurt and other hierarchical theories is whether people can still be morally responsible in cases where their second-order desires are manipulated. For example, let us consider a case where Jones 4 was manipulated to *want to* kill Black. Although his action was not directly manipulated, his will *was*. Our intuition here is that Jones 4's second-order desire was not up to him, and so he cannot be morally responsible for his action. Alfred Mele (1995) provides a widely discussed version of this objection in the case of Ann and Beth. Two agents, Ann and Beth, have similar ambitions, and values. Ann came to develop these motivations through years of self-reflection and experience. Beth, however, came to have her motivations due to brainwashers who have manipulated her to have certain ambitions and values. Consider a case where Ann and Beth both perform a morally wrong action informed by their

background beliefs and reasons. It seems that Ann is morally responsible for this wrong action, while Beth is not. Even though Beth was not directly manipulated to perform the action, her motivations and beliefs were not her own. Mele's case of Ann and Beth shows us that even though they have similar motivations, intuitively it seems that Beth fails to be morally responsible due to the external manipulation of her values and desires.

Frankfurt style cases have paved the way for other compatibilists to develop accounts that focus on the actual sequence of events leading up to an action rather than physically possible alternatives. For example, reasons-responsive compatibilists argue that what is necessary for moral responsibility is not physically possible alternatives, but rather a certain capacity to be sensitive to reasons. John M. Fischer and Mark Ravizza (1998), two reasons-responsive compatibilists, propose a moderate interpretation. A strong interpretation of a reason-responsive view would say that for an agent to be morally responsible, they must always be receptive and responsive to reasons for action. On this strong view, an agent's actions reliably "track reasons" in every case (42). But this standard, Fischer and Ravizza argue, is too demanding for moral responsibility since we cannot hold agents responsible in cases where this "tracking" fails. For example, consider cases where normally rational agents make irrational and morally bad decisions. For example, take two friends who bet twenty dollars on a football game. Adam, who bets on the losing team, gets angry when his team loses and decides to shoot his friend Omar to avoid paying the twenty dollars. This response is obviously irrational and wrong, and thus does not track reason. But we still intuitively believe that Adam is blameworthy.

On their moderate account, Fischer and Ravizza argue that for an agent to be morally responsible they must exhibit a general pattern of being receptive and reactive to reasons. In

other words, if an agent has 1. Generally shown that they can understand at least some reasons for actions (receptivity), 2. Generally shown that they can act on some of these reasons (reactivity), and 3. Came to have these reasons on their own, then they meet the necessary conditions for moral responsibility. Fischer and Ravizza's moderate account allows an agent to be morally responsible in situations where an agent makes a morally bad decision, so long as the agent seems to be "regularly receptive to reasons," (70). However, reasons-responsive accounts face the same manipulation objection that Frankfurt's hierarchical view faces. In cases where an agent lacks control over their receptivity or reactivity, we intuitively feel like an agent cannot be morally responsible. If an agent is manipulated to respond and react to certain reasons, then we intuitively feel like this undermines moral responsibility.

P.F. Strawosn, another compatibilist, (1962) argues that holding an agent morally responsible is an expression of the reactive attitudes. Strawson argues that when we ask about moral responsibility, what we care about is whether other people's actions reflect either positive or negative attitudes towards us. Strawson defines reactive attitudes as "natural human reactions to the good or ill will of indifference of others towards us, as displayed in their attitudes and actions" (67). For example, resentment, gratitude, moral anger, and forgiveness are all reactive attitudes. Strawson's defense of compatibilism is ultimately a defense of reactive attitudes. So, a Strawsonian compatibilist account, and other views like it, do not appeal to the truth of PAP for moral responsibility.

There is a lot more to be said about Strawson and the reactive attitudes. I am only offering a brief exposition here because I will later go into greater detail about the role of the reactive attitudes and how it relates to free will skepticism in Section 3 and in Section 5.

1.2   Libertarian Incompatibilist Views

Two of the main libertarian incompatibilist views are agent-causal views and event-causal views. Once again, these libertarian accounts hold that the truth of determinism would threaten free will and moral responsibility. They all believe that determinism is false and that we have human agency. However, they vary in their accounts of the necessary and sufficient conditions for free, indeterministic actions.

One camp of libertarianism is event-causal libertarianism. Event-causal libertarians argue that a free action is caused indeterministically by a series of events that involve an agent, such as the agent's thoughts, beliefs, desires, or intentions (O' Connor 2011). Robert Kane's (1996) view is one of the most fully developed accounts of event-causal libertarianism. Kane believes that events play a causal role in bringing about an agent's action, but the agent can be ultimately responsible for the action. That is, an agent possesses the kind of control needed for moral responsibility. For an agent to qualify for moral responsibility, either the agent's action is not causally determined, or, if the action is causally determined, then the determined cause (a motive or reason shaped by self-forming actions) must be indeterministically caused. In Kane's view, an agent can be ultimately responsible for a decision that results from their possessing certain character traits, so long as there is a hint of indeterminacy in the agent's history of obtaining those character traits. In any case where an agent is a candidate for moral responsibility, there must be *some* actions that are undetermined, according to Kane. These actions that are not causally determined are called what Kane calls "self-forming actions" which are essentially character forming decisions that an agent makes when they find themselves making a "torn

decision" (74). Self-forming actions indeterministically shape our character, and when our actions stem from this character, we meet the sufficient condition for moral responsibility.

To help illustrate his view, Kane (1999) considers what happens when an agent has to make a torn decision. Here he provides the example of a businesswoman who, on her way to an important meeting at work, witnesses an assault in an alley. On one hand, the businesswoman does not want to be late to her meeting. On the other hand, she knows that it is morally wrong to overlook the incident. She experiences an internal struggle where she debates two courses of action: (1) she can listen to her moral conscience and do something about the assault, or (2) continue on her way to work and pretend to not see anything. Here, Kane wants to emphasize that what the businesswoman chooses is undetermined; both (1) or (2) are physically possible options. The businesswoman must effortfully attempt to overcome her self-interested desires and take the morally correct action, or effortfully attempt to overcome her moral motives to help. The businesswoman's reasons and motives cause her action, and indeterminism is an obstacle that she must overcome either way (226). The businesswoman's contemplation between (1) and (2) constitutes a self-forming action, and she is responsible for whichever option she takes because her decision was undetermined and belongs to her in a significant way. On this view, our self-forming actions accumulate over time and shape our character, motives, and reasons in such a way that allows us to be morally responsible for other actions we take in the future.

Kane discusses the luck objection for event-causal libertarianism. This objection points out that if choices are undetermined, then whether the businesswoman chooses to intervene or not seems to be a matter of luck or chance. If the businesswoman chooses to do the blame-worthy action (1) in World A, but the morally correct action (2) in World B, then her

decision looks like a genuine toss-up. The problem here lies in the lack of contrastive explanation. Assuming there are no differences in her motives and reasons in World A and World B, we can't explain why she chose to perform (1) rather than (2) in World A, but (2) rather than (1) in World B. This lack of explanation makes it seem unfair to blame her in World A and not in World B.

Another main camp of libertarian incompatibilism is the family of agent-causal views. Agent-causal theorists believe that an agent causes the agent's own actions rather than events. Many agent-causal theorists' advocate their view by emphasizing the incompatibilism between *mechanistic* explanations and *purposive* explanations for human actions, as Jaegwon Kim (1998) does. Mechanistic explanations are those that refer to physical and scientific explanations, while purposive explanations are the kinds that appeal to "beliefs, desires, and intentions" (O'Connor 338). In their account of free will and moral responsibility, agent-causal views reintroduce the agent as not only being involved in events, but as an uncaused cause. An agent's possession of causal power gives them the control to settle decisions, and thus effectively bring about their actions. Their internal states can produce a plurality of outcomes, but what settles what will happen is not events, but rather, the agent themself. This causal power, AC libertarians argue, is necessary for moral responsibility.

One agent-causal theorist, Timothy O'Connor (1995) describes his view as such: Any decision that an agent brings about is not entirely due to any sort of reasons, motivations, or environmental factors acting upon them. Rather, the agent brings about an action directly in response to various considerations. As O'Connor puts it, "I am the source of my own activity,

not merely in a relative sense… but fundamentally, in a way not prefigured by what has gone before." (174).

A concern for the AC libertarian is addressing the role of reasons to the agent; some AC libertarians believe that reasons play a causal role, while others believe that reasons can contribute to bringing about an action, but do not play a necessary causal role. Either way, for AC libertarians, our reasons, motivations, and other internal states are insufficient for bringing about an agent's action. The agent themself, above their mental states, is what causes their action. Proposing that an agent causes their own mental states or reasons for action is a seemingly absurd position. When considering the question of what causes the agent, agent-causal theorists must resort to the claim that agents are an uncaused cause which is logically impossible.

1.3   Free Will Skepticism

Free will skeptics are incompatibilists, meaning that they believe that determinism would undermine free will and moral responsibility. What free will skeptic views have in common is that they all seriously consider the possibility that human beings lack free will, or the strongest kind of control required for moral responsibility (Shaw, Caruso, and Pereboom 2019). Some free will skeptics, while either doubting or denying that we have free will, argue that we should maintain our beliefs in free will because abandoning it would undermine morality and many of the other practices I listed above. This leads some, including Saul Smilansky (2000) to argue for an illusionist account where we accept free will skepticism theoretically, but still maintain the illusion of free will and moral responsibility given the devastating consequences associated with relinquishing it. Other skeptical views and the ones that I am specifically concerned in engaging with in this essay include optimistic free will skeptical views. Optimistic free will skeptics such

as Derk Pereboom and Gregg D. Caruso argue that we need not feel dismayed by the fact that we lack free will because giving up the belief that we have free will and moral responsibility would generally be more beneficial overall.

Because humans are never capable of being morally responsible according to the free will skeptic, agents are never truly deserving of blame, praise, punishment, or reward. Many objections to free will skeptical views focus on the pragmatic concerns that arise when considering how we would be able to function without notions of praise and blame, or other concepts based on desert. For example, if we accept free will skepticism, then our ideas about morality, criminal justice, and even our personal relationships may be jeopardized. However, optimistic skeptics argue that we would actually be better off if we relinquished ideas of moral responsibility and accepted the implications of free will skepticism.[5] Additionally, even though all free will skeptics must reject retributivism, or the justification of punishment grounded in desert, free will skeptics differ in what we should do about it.

If humans always lack the kind of control required for free will, then this calls into question many of our everyday practices. These practices include our motivations and meaning in life, the notion of morality itself, reactive attitudes expressed in genuine interpersonal relationships, and our judicial and legal systems that presuppose the legitimacy of free will and basic desert, just to name a few. In what follows I will focus primarily on Derk Pereboom and Gregg D. Caruso's varieties of optimistic free will skepticism.

---

[5] Caruso, Gregg, "Skepticism About Moral Responsibility", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition)

2. Hard Incompatibilism

Derk Pereboom (2001, 2004) argues that regardless of the truth of determinism, we never have the control required for moral responsibility. Pereboom argues

1. We fail to be morally responsible if compatibilism is true

2. We fail to be morally responsible if event causal libertarianism is true

3. Agent-causal libertarianism is naturalistically implausible

   Therefore

4. We ought to take seriously the idea that we lack free will and cannot be morally responsible

Pereboom argues that regardless of the truth of determinism we never have the control required for moral responsibility and describes this position as hard incompatibilism (2011).

Against compatibilism, Pereboom offers his famous four-case manipulation argument to show that we always lack the kind of causal control required for free will if compatibilism is true (74-82). Pereboom's (2001, 89-126) four cases depict Professor Plum who decides to murder White for selfish reasons, and Professor Plum succeeds in all cases. All cases meet the compatibilist requirements for moral responsibility including but not limited to: the fact that the action is not out of character for Plum, Plum desires to kill White, Plum has the second-order volition to kill White (or wills to kill White, and wants to will to kill White), and Plum is sensitive to moral reasons not to do so. Pereboom presents the cases as follows:

Case 1: Plum was created by neuroscientists who can and do directly manipulate his mental states, attitudes, and actions at any given moment, but he is as human as he possibly can be, given this history. He is rationally egoistic thanks to the neuroscientists' intervention.

Case 2: Plum is like an ordinary human being, but he was created by neuroscientists who cannot directly control him, but nonetheless programmed him to be rationally egoistic at birth such that he wishes to desire to kill White.

Case 3: Plum is an ordinary human being, except that he is determined by the rigorous training practices of his home and community in such a way that causes him to wish to desire to kill White.

Case 4: Determinism is true, and Plum is an ordinary human being. Plum's killing White results from his reasons-responsive process of deliberation. His egoistic reasons are very powerful, and these reasons combined with background circumstances deterministically result in Plum murdering White (74-82).

Cases 1-4 are supposed to show different ways that Plum is causally determined to kill White. In Case 1, intuitively we feel that Plum is obviously not morally responsible for killing White. This generalization strategy that Pereboom employs demonstrates that there is no principled difference in the degree of control that Plum has in Cases 1, 2, and 3, to Case 4. It seems that if one concedes that Plum is not morally responsible in Case 1, he can't be morally responsible in the following cases because there are no morally relevant differences between the cases (intuitively). Case 4 is how we would expect the world to be if determinism were true; If Plum is not morally responsible in Case 4, then none of us are morally responsible if determinism is true. If we take the generalization strategy to be convincing, then compatibilism is implausible. Considering this generalization strategy, compatibilism seems implausible.

Against event-causal libertarianism, Pereboom argues that the conjunction of the luck objection and the disappearing agent objection render the view incoherent. Because event-causal

libertarians believe that agents are caused indeterministically and solely by prior events, critics question where in the process does the agent settle the action? On event-causal libertarianism, it seems like the agent plays an insignificant role in the process, if any role at all. This disappearing agent objection captures the lack of a crucial connection between the agent and their action that proponents of a source view require for moral responsibility. Pereboom also highlights the luck objection that I discuss in Section 1. These two objections, the disappearing agent objection and the luck objection, taken together highlight how it is difficult to see the agent having the relevant control necessary for moral responsibility on an event-causal view.

Turning to agent-causal libertarianism, Pereboom argues that the agent-causal view is that it just is not plausible given our best physical theories (Pereboom 2014, 66). Our best physical theories tell us that the physical world is law governed. According to the agent-causal view, agents are supposed to be substances with efficacious agent-causal abilities. On the agent-causal view, an agent is not caused by external facts about the physical world, or by events such as mental states. We would expect that an agent, possessing its agent causal power, would at some point diverge from the physical mechanistic laws including psychological and neuroscientific laws stemming from, say, our brains. However, it seems that agents often cohere fortuitously with psychological and other physical rules that govern our world. One could argue that agents, even possessing their causal powers, just happen to always fall in line with the psychological and physical laws governing our universe, even though they are not causally determined by them. But, as Pereboom (2014) points out, "this proposal would… involve coincidence too wild to be credible" (66). In order to avoid this objection, often called the wild-coincidence objection, the agent-causal libertarian must offer an account of how we could possibly accurately predict what

action an agent will take, given all relevant information about their mental states, environmental conditions, attitudes, and other factors that play a role in the action they take. The idea that physical facts about our universe could accurately predict what action an agent will take, on both the deterministic or indeterministic view then it appears that an agent's action is no longer free, and agent-causation looks less plausible.

Given that there is (1) a significant theoretical worry for compatibilism (shown through Pereboom's four-case-manipulation argument), (2) a daunting luck and disappearing agent objection for event-causal libertarianism, and (3) the fact that our best physical theories do not support the agent-causal view, we arrive at the conclusion that we *always* lack the kind of free will that moral responsibility requires. This argument for hard incompatibilism is Pereboom's negative argument, or his argument against the other theories on free will.

Pereboom and other optimistic free will skeptics argue that many of our practices that presuppose free will and moral responsibility can be and often are harmful. Given this fact, we should not be dismayed by the proposition that we lack free will and cannot be morally responsible, and moreover, that we have reason to be *optimistic* considering this fact. In the next section I will discuss how Pereboom motivates his argument for the positive implications of free will skepticism, starting with the implications for a set of attitudes that seem essential to interpersonal relationships - the reactive attitudes.

3. Reactive Attitudes

P.F. Strawson (1962) argues that "we demand some degree of goodwill or regard on the part of those who stand in these [personal] relationships to us." Strawson calls these "reactive attitudes," defining them as "naturally human reactions to the good or ill will or difference of

others towards us, as displayed in *their* attitudes and actions," (53) We care about the intentions of others and whether they hold esteem, indifference, or contempt towards us. Strawson posits that these attitudes are important to us because they acknowledge "human self-respect and its connection with the recognition of the individual's dignity" (49). These attitudes include gratitude, love, forgiveness, indignation, and resentment to name a few.

In Strawson's view, what is necessary for a person to be morally responsible is that they must be a proper target of the reactive attitudes toward themselves and towards others. Strawson argues that some of these attitudes are also crucial for our interpersonal relationships, and that to take them away would make human life seemingly bleak and meaningless. What would our lives look like without expressions of love, gratitude, and forgiveness? Other reactive attitudes, including blame, anger, and resentment are reactions to injury or indifference and are important because when we express them, we are acknowledging other people's and our own attitudes and intentions. I will explore this claim in greater depth in Section 5. Strawson argues that these attitudes are essential not only because they are natural expressions of our moral attitudes, but also because of their efficacy in regulating behavior in desirable ways. Pereboom argues that not only should we accept free will skepticism theoretically, but also that we should accept its normative implications. If we take free will skepticism seriously and nobody is ever free and responsible, then we must question our expression of reactive attitudes. The reason for this is because our reactive attitudes respond to other people's attitudes and intentions towards us. If people lack free will, then they cannot control their attitudes or intentions. Given the weight of free will skepticism, we have reason to question some of our practices that involve blame, resentment, and moral anger.

One common objection to the free will skeptic is that a world where people do not express reactive attitudes at all is bleak and meaningless. However, Pereboom points out that not all our attitudes are threatened by free will skepticism. There are reactive attitudes that we can and should keep as hard incompatibilists. These reactive attitudes are those attitudes that are not harmful, and moreover, can be beneficial regardless of whether we deserve them. Some examples of these include forgiveness, gratitude, and love. Pereboom agrees that some of these attitudes are important to our interpersonal relationships, or at least meaningful ones.

One might argue that if we accept free will skepticism, forgiveness is undermined because it involves acknowledging that a person has done something blame-worthy, but we choose to withhold or give up blame and resentment towards them for their transgression. However, this is just one view of what forgiveness is. We can think of many other different cases where it seems like forgiveness does not involve resentment or blame. Hieronymi (2001) argues that "If I ask for forgiveness, I am… asking you to believe me when I say that I no longer see what I did to you as acceptable, to recognize, and so ratify my change of heart" (554). Pereboom also points to forgiveness in some cases as "willingness to cese to regard past wrongful behavior as a reason to weaken or dissolve a relationship" (189). Pereboom argues that these conceptions of forgiveness are not inconsistent with free will skeptical views and so we have reason to keep the attitude of forgiveness (2001).

Pereboom also argues that can keep other non-reactive attitudes including gratitude and love. One might worry that if someone is grateful towards an agent, then that agent is "morally responsible in the basic desert sense for a beneficial act" (189). Gratitude often involves notions of praiseworthiness, which could arguably be undermined by the skeptical view. However,

Pereboom argues that we can and are still grateful to people in other cases where we do not think they are not morally responsible. For example, we feel grateful towards children who show us kindness. We feel gratitude towards our loved ones often not because we think they deserve it, but simply because we enjoy having meaningful relationships with them. Pereboom also highlights the fact that "gratitude involves joy as a response to what someone has done," and nothing about the skeptical view undermines this (190).

Like gratitude, one might worry about love because we often express love for one's moral character. Love might be threatened by free will skepticism because it suggests that people lack control over who they are in an important sense. However, parents do not love children for their morally praiseworthy character, and children don't love their parents for this reason either. While one *can* love a person for their moral character, it is often not one's moral character alone that causes someone to love a person. Pereboom argues that love fundamentally involves "wishing well for the other, taking on aims and projects of the other as one's own, and a desire to be together with the other" (190). He points out that free will skepticism does not threaten this.

Pereboom argues that there are certain valuable attitudes such as forgiveness, gratitude, and love that do presuppose desert, and thus, we can maintain these attitudes under a free will skeptical view. However, there are other kinds of attitudes that Pereboom believes we should treat differently than forgiveness, gratitude, and love. In addition to these attitudes, there are reactive attitudes such as resentment and indignation that Pereboom argues are irrational and harmful. Resentment refers to the "anger with an agent due to a wrong he has done to oneself," and indignation is "anger with an agent because of a wrong he has done to a third party" (179). Resentment and indignation are different because they invoke a "strong conviction of moral

responsibility in the basic desert sense [and] involve especially malevolent harm done to another. Pereboom proposes that we should replace these harmful attitudes with analogues that do not presuppose free will and desert.

Pereboom argues that we can and should replace resentment and indignation with disappointment, hurt, and sadness. Pereboom argues that these analogues can do similar work to resentment and indignation in conveying our feelings when people wrong us, but unlike their reactive counterparts they do not presuppose basic desert. When we communicate disappointment, hurt, and sadness, we are not necessarily assuming that people could have done differently and that they are morally responsible for what they did. Consider a case with two life-long friends, Bill and Candace. Bill's partner cheats on Bill with Candace. When Bill finds out, he expresses disappointment and hurt at the situation, but not resentment. Candace feels terrible and genuinely sorry for betraying Bill's trust. Pereboom would cite this as a case where we can see how alternatives to resentment do the same work as resentment, but are not attached to desert.

But if we accept the idea that moral anger is harmful overall and that we should give them up, then we can at least mitigate resentment and indignation in a broad sense where we are able to assess actions from a non-emotionally fueled standpoint. He argues that this would be beneficial for us in our personal relationships and in other practices that presuppose free will and basic desert. If we do not express anger, resentment, or indignation toward our friends and loved ones, we would not be subjecting them to undeserved harm and would thus expect our relationships to improve. Pereboom argues that "we might be able to take the measures that

would moderate or eliminate resentment and indignation, and in particular their overt expressions," and that we would do so "for the sake of morality and rationality" (2014, 152).

Pereboom argues that by expressing moral anger, we are being irrational and unfair to people who are on the receiving end of our expression of these attitudes. Pereboom asserts it is morally wrong to subject someone to this attitude given that it can cause emotional and physical pain and is undeserved. Further, he points out that these attitudes "can give rise to destructive resistance instead of reconciliation" (181). For example, take violent protests as opposed to non-violent protests. Violent protests can be seen in the form of looting, burning buildings, and other destructive acts. Resistance motivated by anger can make those not participating in the protests view the violent activists as unwilling to negotiate and unlawful, thus creating further division.

While there are some reactive attitudes such as indignation and resentment that can be replaced with non-reactive analogues, there are other reactive attitudes that do not have apparent analogues (such as moral anger). Pereboom acknowledges that we can still reasonably expect that expressions of reactive attitude would not be eliminated, entirely. Our reactive attitudes are deeply ingrained in our everyday practices. When our friends break their promises to us or our partners undermine our trust in them, we express that we are angry at them and intuitively think we are justified in doing so. Often, we can't help our expressions of moral anger, but if we accept free will skepticism normatively speaking, we can work on attempting to eliminate it in a broad sense. Despite this, Pereboom argues that free will skepticism recommends that we give this anger up because it is based on false beliefs, and a retraction of anger is in many cases is will also yield more positive results.

In summation, Pereboom argues that under free will skepticism we have three options regarding our current attitudes. Some attitudes (including forgiveness, gratitude, and love) are not threatened, and we can maintain them. For attitudes such as indignation and resentment we can replace them with analogues that do not presuppose desert in the same way that their reactive counterparts do. Finally, some reactive attitudes, specifically moral anger, we would be better off eliminating altogether even if we cannot actually do so.

4. Theories of Criminal Punishment

Accepting the truth of free will skepticism has significant implications for not only reactive attitudes, but also for our methods of punishment and criminality in our judicial system.[6] Free will skeptical theories face serious moral challenges when it comes to accounting for the justification for criminal punishment. Pereboom importantly notes that "punishment inflicts harm, and in general, justification for harm must meet a high epistemic standard" (Pereboom 2014, 158). This high epistemic standard requires that we should have strong evidence that people are deserving of punishment, especially given its serious consequences and harms. Given how Pereboom motivates the negative argument for hard incompatibilism, we have reason to doubt that anyone is ever deserving of punishment. If no one ever has free will or the control necessary for moral responsibility, then we must question justifications for punishment that appeal to basic desert. Retributivism holds that "punishment of wrongdoers is justified for the reason that they deserve pain, deprivation, or death just because they have knowingly done wrong" (Shaw, Pereboom, & Caruso 2019). Free will skepticism requires that we should abandon retributivism as a justification for punishment because if people are never deserving of

---

[6] When referring to criminal punishment and the judicial system, I am specifically addressing the U.S.'s judicial system, as it is the one I am most familiar with, but globally many other countries presuppose retributivism in their treatment of criminals as well.

blame, then they are never deserving of punishment either (Pereboom 2014). This section

focuses on a couple of different non-retributive theories for punishment that free will skeptics

have put forward including deterrence theories, moral education theories, and self-defense

theories.

According to deterrence theories of punishment, we can justify criminal punishment on

the grounds that it prevents other potential criminals from committing crimes. These theories,

called deterrence theories, often end up falling under utilitarian views, or views that advocate for

acting based on the principle of maximizing utility by examining the benefits and the drawbacks

of an action. Jeremy Bentham's (1823) view prominently features the classic deterrence theory.

On Bentham's utilitarian view, the state should adopt the policy of criminal punishment that aims

to maximize utility. The state is only justified in punishing criminals if doing so maximizes

utility. Deterrence theories may seem appealing to free will skeptics given that they do not appeal

to desert or rest on the assumption that we are free and responsible agents. However, many have

raised threatening moral objections against utilitarian theories like Bentham's.

First, utilitarian deterrence theories justify punishments that are both intuitively severe

and morally wrong. We can see this when we consider what actions utilitarian deterrence theories

permit us to take. If the utilitarian goal is to maximize utility, then we need not care about the

means used to achieve these ends; utilitarian recommendations often fail to "conform to our

intuitions about fairness" (Pereboom 2014, 164). McCloskey's (1965) objection highlights this

problem, arguing that deterrence theories justify punishing the innocent. If punishing the

innocent maximizes utility and potentially deters future crime, then the utilitarian would

recommend that we should punish the innocent so long as it will yield the best results.[7] But, we intuitively think that those who did nothing wrong should not be punished, so utilitarian deterrence theories give us recommendations that are intuitively morally wrong.

One further daunting objection against utilitarian deterrence theories is the *use objection*. The use objection centers around our moral intuition that it is wrong to use people as a means for our own safety, as this violates the deontological principle of treating people with dignity and respect. Pereboom points out that a problem for utilitarianism is that it "sometimes requires people to be harmed severely, without their consent, in order to benefit others" (165). This is obviously true in cases where we punish criminals in order to motivate law-abiding behavior among other citizens. Under deterrence theories, we are using criminals as instruments to achieve a certain outcome, undermining their worth as human beings. So even though hard incompatibilists can appeal to deterrence justifications for punishment, lingering moral worries should discourage them from doing so.

Since deterrence and utilitarian theories fail as strong non-retributive candidates, we now turn to moral education theories. Moral education theories justify punishment of criminals to morally educate them, and thus improve their future behavior. Such theories often draw on the analogy of punishing children. We often punish children not because we believe they deserve it, but rather to educate them about what is morally right and wrong. Herbert Morris's moral education theory offers the example of a parent who deprives their child of the reward of a game after the child cheats. The withdrawal of the game makes it clear to the child that there are unpleasant consequences associated with bad behaviors, such as cheating (Morris 1981). Morris

---

[7] McCloskey, H. J. (1965). A Non-Utilitarian Approach to Punishment. *Inquiry: An Interdisciplinary Journal of Philosophy* 8 (1-4):249 – 263.

also argues that punishment can communicate to a child the seriousness of a moral offense, and varying levels of punishment can help illustrate important moral boundaries. We generally think that punishing a child is crucial in their moral training and can see how a complete absence of punishment and discipline can end up committing a serious disservice to the child as they attempt to navigate the world without proper knowledge of moral rights and wrongs. Moral education theories do not depend on desert; these theories do not presuppose that the child deserves the punishment, meaning hard incompatibilists can appeal to these theories

One potential worry for this view is that we lack compelling evidence that suggests that punishment can help morally improve adults, particularly adult criminals, the same way that it improves children. Pereboom points out that unlike children, adult criminals typically understand their society's accepted moral code (161). While there are some adult criminals who may genuinely not understand that their actions are morally wrong (like many psychopaths for example), most criminal adults understand what constitutes a moral right or wrong. Further, children lack the experience in moral training that adults typically have, which is a reason that we often refrain from harshly punishing children, especially when their offense is slight. Pereboom (2014) also highlights the fact that children are more psychologically malleable and are punished within the context of a caring and interpersonal environment. Ultimately, it seems that there are too many morally relevant differences between the state's treatment of an adult criminal and a parent's treatment of their child. If punishment is to be justified by appealing to moral education theories, then a greater deal of empirical evidence is needed that demonstrates that moral education in the context of criminal punishment actually works (Pereboom 2014).

Thus, while theoretically there is nothing about moral education theories that is inconsistent with hard incompatibilism, it still seems like we should not accept them.

Finally, some free will skeptical accounts try to justify punishment by appealing to people's general right to self-defense and the defense of others. Daniel Farrel (1985) argues that each of us have a right to self-defense. This is an uncontroversial claim. If Maeve randomly charges at Otis in public with a knife in her hand, it is not unreasonable for Otis to react first and knock out Maeve. We can intuitively see how the right to self-defense overwhelms the use objection here. In cases where someone is being directly threatened, that person has the right to harm an unjust aggressor a reasonable amount to prevent the aggressor from harming them. Similarly, Farrel argues that we can apply this to the state in the context of punishment. This right to self-defense is a right that the state has a duty to protect given that the state is obligated to protect our physical safety from violence. Farrell claims that the state, acting as a proxy for us, is justified in carrying out punishment against unjust aggressors to prevent them from harming us. These claims together bring us to the conclusion that the right to self-defense could justify state punishment. Self-defense theories do not appeal to desert, so hard incompatibilists could use them.

The right to self-defense is intuitive in cases where a stranger is unjustifiably trying to attack someone. However, one objection to self-defense theories is that they fail to compellingly motivate the connection between a person defending themself and the state being justified in punishing people. There are important morally relevant differences between the cases. The state does not only protect you from the moment of attack to secure your safety, but it also determines what happens to the criminal long after the direct threat has been removed. While it is

understandable for Otis to knock Maeve out when she attacks, it would be questionable for Otis to continue to harm or punish Maeve afterwards. The state continues to punish and harm criminals after they commit crimes, even when it is unclear whether they still pose a direct threat.

So far, we have canvassed different theories that hard incompatibilists could justify criminal punishment including deterrence theories, utilitarian theories, moral education theories, and self-defense theories. However, the normative, moral, and empirical consequences make all of these options look untenable. This motivates a need for some new alternative. Pereboom and Caruso argue that the best alternative is a quarantine theory for punishment that is fundamentally grounded in self-defense but makes sure to account for the problematic implications of these kinds of theories.

5. Quarantine Theories

Pereboom believes that out of the theories of punishment left open to free will skeptics, invoking the right to self-defense is most intuitive and faces the least challenges. This section focuses on how Pereboom draws upon self-defense theories in order to develop his own account where he argues that we are justified in quarantining criminals to remove threats they pose. In addition to Pereboom's quarantine theory, I will also discuss Caruso's public-health model which expands upon Pereboom's quarantine theory.

5.1 Pereboom's Quarantine Theory

Pereboom's (2013) own account of criminal punishment invokes the right to self-defense, which does not justify intentional harm, but instead justifies incapacitation of aggressors. To support this claim, Pereboom provides an analogy:

"Suppose that someone clearly aims to kill you, and that to prevent his doing so you may knock

him out with a baseball bat...suppose that he trips over the toys on the floor, and this allows you

to pin him to the ground and tie him up. At this point is it still legitimate for you to knock him out

with the bat? To do so would not be justified by the right to harm in self-defense." (168)

The same line of thought applies when you attempt to secure other's safety, as well. If the

aggressor no longer poses a threat, then invoking the right to self-defense seems to best justify

incapacitation, not necessarily punishment. Returning to the broader discussion about the state's

role, the state cannot punish aggressors who are already in custody and who no longer pose a

threat. Given that people have the right to self-defense, the state is only legitimately justified in

incapacitating criminals so that they cannot harm other people.

Pereboom's view is justified when we consider the legitimate functions of the state within

the context of freedom from harm which include "protecting its citizens from significant harm

and providing a framework for human interaction to proceed without significant impairment,"

(159). These basic roles, Pereboom claims, do not necessarily presuppose basic desert

themselves. Pereboom ultimately justifies his own view's incapacitation of criminals by invoking

only the right to harm in self-defense and defense of others from dangerous criminals and

believes that punishment would only be justified in cases where it is necessary for protecting

oneself and others.

Pereboom proposes that we adopt an incapacitation model, whereby we incapacitate

dangerous criminals the same way that we treat carriers of dangerous diseases. If we adopt

Pereboom's view, we are not permitted to inflict more harm than what is required for removing

the threat that criminals pose. Pereboom argues that by drawing this analogy encourages us to

focus on "the rehabilitation and well-being of the criminal that would alter much of the current

practice" (156). This account only permits the minimal harm required for securing people's

protection, and moreover, allows for endorsement of preventive measures of reducing crime,

including "providing for adequate education and mental health care," (156) or addressing other

systemic issues that cause people to resort to crime in the first place.

Pereboom acknowledges that the use objection that poses such a problem for

utilitarianism still holds much force and shouldn't be dismissed, even though Pereboom and

Caruso believe otherwise:

> "We have always maintained the *principle of least infringement,* which holds that the least
>
> restrictive measures should be taken to protect public health and safety. This ensures that criminal
>
> sanctions will be proportionate to the danger posed by an individual, and any sanctions will be
>
> proportionate to the danger posed by an individual, and any sanctions that exceed this upper
>
> bound will be unjustified." (Caruso 2019, 61)

Even though Pereboom emphasizes that his view aligns with the *principle of least*

infringement, it is still the case that preventative detention of criminals to secure our own safety

still violates an important deontological principle. Even if we avoid overly invasive and harmful

punishment of criminals in incapacitation, we are still denying people the opportunity to live a

fulfilling life and the opportunity to pursue other meaningful opportunities. Pereboom

acknowledges that he still needs to address the use objection, but essentially tries to lessen its

force by prioritizing humane treatment of criminals through rehabilitation and their reintegration

back into society. Given that preventative detention would likely cause harm to dangerous

criminals, Pereboom points out that we should only be allowed to incapacitate criminals that are

most dangerous, just like we would only quarantine people with the most dangerous diseases.

While the use objection still holds considerable force, at the end of the day, the incapacitation

account does better than its alternatives; it does not accept excess punishment or condone punishment of the innocent like deterrence and utilitarian theories do.

Given that we should liken the treatment of criminals to the treatment of carriers of dangerous diseases, then punishing criminals more severely than what is needed to remove the threat they pose is unjustified. So, excessive punishment is not justified on the view. Further, the degree of invasiveness of our preventive detention measures for carriers of dangerous diseases varies depending on the seriousness of the disease and the threat of contagion. Similarly, the degree of invasiveness in our incapacitation of criminals should vary depending on the threat that the criminal poses, which includes the severity of the criminal's crime. This forward-looking view also not only allows, but also encourages rehabilitating criminals back into society and taking preventive measures against criminality in the first place.

5.2 Caruso's Public Health-Quarantine Model

Another optimistic free will skeptic, Gregg D. Caruso (2019), defends a view that draws on Pereboom's quarantine model, but further develops it "within a broader justificatory framework drawn from public health ethics" (60). Caruso motivates his view with the following argument (60):

(1) Free will skepticism maintains that criminals lack moral responsibility in the basic desert sense

(2) Carriers of dangerous diseases are not responsible for having contracted their disease

(3) Generally, we agree that it is acceptable to quarantine these carriers of dangerous diseases to protect oneself and protect others.

(4) Similarly, we are justified in incapacitating criminals to protect oneself and others, even though they lack moral responsibility.

Caruso agrees with Pereboom that utilitarian and consequentialist arguments that could advocate for a quarantine-like model should be resisted due to their problematic implications, including problems with punishing the innocent, overly harsh punishment, and using people as means. He believes that the use objection and the possibility of harming the innocent hold especially significant weight and should discourage us from adopting such views. Like Pereboom, Caruso believes that excessive unnecessary penalties against criminals are unjustified. Caruso also highlights the fact that the incapacitation account and its focus on fairness demands the criminal's rehabilitation, well-being, and reintegration into society.

Caruso's public health-quarantine model offers a few further contributions to Pereboom's incapacitation model. An interesting additional feature of Caruso's public health-quarantine model is that it requires that we take proactive measures to prevent such outbreaks from occurring. That is, in Caruso's view, the criminal justice system would not only focus on deciding what to do after people commit crimes, it would also primarily function to prevent criminality in the first place. Caruso argues that just as there are social determinants of health (SDH), we should recognize that there are also social determinants of criminal behavior (SDCB). Importantly, Caruso's view recognizes how environmental conditions such as pre-existing medical conditions, homelessness, and poverty can all lead a person to commit crimes. The public health framework that Caruso advocates prioritizes social justice as a "foundational cornerstone to public health and safety" (62). Caruso advocates for an incapacitation model that is grounded in a theory of justice, favoring a "capability approach" (63). This means that the

model would see human well-being as the proper end. Caruso, drawing on work of other capability theorists, argues that social justice should be grounded in six key features of human well-being: health, reasoning, self-determination, attachment, personal security, and respect (Powers and Faden 2006, Caruso 2016, 2017). Caruso's view of justice means that the state's role is to sufficiently ensure that it prioritizes human well-being using these six measures.

On the capability approach, state institutions should "aim to expand people's capacities," or "their freedom to promote or achieve *functionings* that are important to them" (Caruso 2019, 63). These functionings include having good health, safety, mobility, security, and other meaningful life states that constitute human well-being. According to Caruso, the state has an obligation to prioritize justice, and given the capability approach, this means that the state has a legitimate role to promote genuine opportunities for its citizens to fulfill the six dimensions of human well-being. If a state's public health institutions fail to secure the population's safety and ensure sufficient health, then we consider it a grave injustice. Caruso's model applies the same logic to criminality; if a state's social institutions limit people's opportunities, and thus their capabilities, and this causes people to turn to crime, then the state has committed a grave injustice against those people. So, if we were to adopt Caruso's public health-quarantine model, we would do so to not only promote public health and prevent crime, but also for the sake of justice and fairness.

6. Objections to Optimistic Free Will Skepticism

In this final section, I will focus on major objections to features of Pereboom's and Caruso's optimistic free will skeptic views. First, I will raise some objections against Pereboom's

arguments for the reactive attitudes. I then move to some moral concerns regarding Pereboom's incapacitation model and Caruso's public health-quarantine model.

6.1 Against Analogues for the Reactive Attitudes

Pereboom (2014) argues that, given free will skepticism, we should give up some of our reactive attitudes, particularly those that he believes are harmful and presuppose basic desert. Pereboom argues that we should get rid of moral anger, resentment, and indignation, and instead replace these reactive attitudes with disappointment, hurt, or sadness.

Pereboom acknowledges the value that these angry reactive attitudes have, specifically bringing in Shaun Nichols' (2007) argument. Nichols argues that Pereboom's analogues, especially moral sadness, are inadequate substitutes for moral anger. Nichols argues that there is a breadth of empirical evidence that shows that reactive attitudes, such as resentment and indignation discourages cheating, mistreatment, and other harmful behaviors. However, other studies (Lazarus 1991) show that for infants, sadness tends to produce no action tendency whatsoever. We can draw the conclusion, then, that sadness typically causes people to give up. Moral anger, however, tends to motivate action, which is what injustice often demands. We can see how this is true when we observe historical social justice movements, such as the Civil Rights Movement in the U.S. and the global efforts against Apartheid in South Africa in the 1960s. Today, what has given the Black Lives Matter campaign, LGBTQ+ movements, and other social justice efforts so much weight is the moral condemnation and anger invoked by injustice. What sprung these movements into action and continue to inspire continuing battles for marginalized groups' rights, are reactive attitudes that Pereboom argues we relinquish.

Pereboom concedes the historical and ongoing importance of the reactive attitudes in motivating resistance against oppression. He points out that social justice movements have been crucial in helping marginalized groups gain freedom and equality. Even if we concede Nichols's argument about the usefulness of moral anger, Pereboom says that moral anger still has more drawbacks than benefits. Especially in our everyday interactions with people, most notably those who we are close to, our expression of resentment, indignation, and anger all cause undeserved harm. Moreover, Pereboom also points out that historically some social justice leaders, including Martin Luther King Jr. and Mahatma Gandhi, have had great success at resisting injustice without being fueled primarily by resentment or indignation (Nelkin 2011).

However, it is seriously questionable that these movements would have made any tangible progress without people who *were* motivated by moral anger and resentment. Further, resistance motivated by moral anger does not always equate to violent resistance. In fact, in most cases it does not. Martin Luther King Jr. and Mahatma Gandhi were famous for peacefully protesting, but we do not know whether they were morally angry or not. Consider Maya Angelou, an American poet and civil rights activist. On injustice, Angelou argued,

> "You should be angry… So use that anger. You write it. You paint it. You dance it. You march it. You vote it. You do everything about it. You talk it. Never stop talking it" (Angelou 2006).

I believe Angelou's quote illustrates many ways that we can express moral anger that are non-violent. I do not find Pereboom's response to the value of moral anger in social justice movements to be persuasive. His view still ultimately argues that no moral anger is justified, even if it helps people resist injustice, which would be fatal given that there is still much inequality and oppression.

What of replacing some of the reactive attitudes with analogues in personal relationships? Pereboom still believes that we would generally benefit *overall* from these substitutions in relationships. Our expressions of blame, resentment, and anger are harmful, especially in our interpersonal relationships. These reactive attitudes can cause physical and emotional harm and are often destructive. Pereboom's argument only works if he is correct in assuming that the analogues are less harmful than their reactive counterparts, but we should not accept this assumption so easily.

Take Pereboom's own example of a parent whose child does something harmful to another person. In Pereboom's view, the parent should choose to express moral disappointment in their child rather than indignation. However, the psychological harm caused by a parent's disappointment often feels *worse* than harms caused by a parent's anger. Parents may be disappointed when their child fails to live up to their parent's expectations. Moving beyond young children, parents may express disappointment in their adult children, as well. A parent may express disappointment if their child fails to meet their expectations in career success or in relationships.[8] One comprehensive study exploring Chinese adults finds that the correlation between parental perceptions of their children's success and problems and their adult children's psychological well-being are statistically significant.[9] The results for this study show that Chinese children attach great importance to their parents' psychological feelings about their own success and problems, and this can significantly impact a child's well-being. Caring about

---

[8] Cichy, K. E., Lefkowitz, E. S., Davis, E. M., & Fingerman, K. L. (2013). "You are such a disappointment!": negative emotions and parents' perceptions of adult children's lack of success. *The journals of gerontology. Series B, Psychological sciences and social sciences*, *68*(6), 893–901.

[9] Yang, J., & Zheng, Y. (2019). Links Between Perceptions of Successes, Problems and Health Outcomes Among Adult Chinese Children: The Mediating Role of Perceptions of Parents' Feelings and Intergenerational Relationships. *Frontiers in psychology*, *10*, 2551.

parental perceptions is not particularly unique to Chinese culture; I am sure many people who have ever been on the receiving end of their parents' disappointment understand the guilt we feel when we let our parents down. Causing another person's sorrow, disappointment, and pain could arguably cause greater damage than their reactive counterparts to the receiver of these kinds of attitudes. My argument here is not that parents should never express disappointment in their children, but rather that Pereboom too quickly assumes that analogues like disappointment inflict less harm than resentment and indignation.

Moving away from the example of parents' disappointment in their child, we should also consider whether disappointment is an appropriate substitute for anger in interpersonal relationships between two morally educated adults. We may think that disappointment is an appropriate reaction for a parent to express towards their child. However, when a moral agent expresses disappointment towards another moral agent, it appears condescending. Consider typical cases where we believe disappointment to be appropriate: a parent's disappointment in their child, a teacher's disappointment with a student, or a coach's disappointment with their star player. In all of these cases, there is a clear power imbalance at play that makes disappointment seem appropriate.

Consider a case with two coworkers. Alex and Lindsey work together at a coffee shop. Lindsey asks Alex to cover a shift for her because at the last minute she realized that she was extremely unprepared for an exam coming up the following week. Alex agrees to cover Lindsey's shift, but then forgets about it and decides to attend a party during the time where she was supposed to take Lindsey's shift. Lindsey is fired because their manager is very strict and believes that Lindsey did not show up for the shift that she was scheduled for. Lindsey later finds

Alex and tells her that she is not mad at her friend, but rather that she is disappointed in her. Lindsey and Alex both started this job at the same time, and neither are positioned higher in the company than each other. In this case, even though Alex failed to fulfill her promise to Lindsey, she gets annoyed at Lindsey's disappointment. Alex thinks, "who does this girl think she is? I understand if she were mad at me, but for her to express disappointment at me gets on my nerves." This case illustrates how disappointment fails as a substitute for anger because it is patronizing, especially in a relationship between two equals.

Let us imagine another case where an agent would express disappointment in their friend rather than resentment or indignation. Amy and Eric are taking a class on moral psychology together. On an exam, Amy failed to study the night before and resorts to looking at Eric's exam and starts copying his answers. The professor notices but misunderstands the situation and believes that Amy attempted to cheat by looking at Eric's answer sheet. After class, the professor pulls aside both Amy and Eric. Eric profusely denies that he tried to cheat and desperately looks at Amy to vouch for him. However, Amy does not say anything, and the professor assumes that Eric did in fact attempt to cheat on the exam by looking at Amy's answer sheet. Eric receives an F on the exam and is referred to the dean's office for academic misconduct. Is it appropriate in this situation for Eric to express disappointment in Amy? We intuitively think not, but rather that resentment is more appropriate. Eric perceives that his friend wronged him, and Eric suffered serious consequences due to his friend's actions. An attitude such as disappointment seems unsatisfactory to accurately capture how Eric should perceive Amy's transgression. Moreover, if Eric expressed disappointment in Amy, it would appear condescending in this case as well.

These kinds of cases demonstrate why Pereboom's analogues intuitively seem inadequate. Strawson (1962) also argues that the expression of moral anger itself communicates that the angry person views a wrongdoer as a person who belongs in the moral community. Essentially, the expression of reactive attitudes shows that we are emotionally engaged with people on a personal level, and this cannot be replaced with a non-reactive attitude like disappointment. Moreover, our expression of resentment and indignation towards others shows that we respect them in an important sense.

So, Pereboom's argument that disappointment, hurt, and sadness are less harmful than resentment and indignation is not one that we should easily accept. We can see how this is true when we look at the harms of parents' expressions of disappointment in their children. We should also doubt that these analogues can act as sufficient replacements when we consider cases when one person expresses disappointment rather than resentment with their equal. In these cases, disappointment is condescending and inappropriate towards its target. Further, by expressing anger with someone, we are indirectly viewing them as a person who belongs to the moral community. For these reasons we should doubt that there are as many benefits to Pereboom's analogues as he claims.

6.2 Injustice and Quarantine Theories

Pereboom uses a quarantine analogy to advance his incapacitation model, arguing that we should treat dangerous criminals the way we treat carriers of dangerous diseases. The public health model captures a very important feature of our justice system, namely that it focuses too much on retributivism and instead that it would be more just and beneficial to prevent crime from happening in the first place. Caruso's model highlights what our justice system fails to

acknowledge, namely that racism, sexism, poverty, and other systemic issues disadvantage those who are more likely to encounter our justice system and thus, public health more broadly. However, this section is dedicated to pointing out some concerns that arise if we fully embrace this view.

First, Saul Smilansky (2000) highlights how the incapacitation account fails to sufficiently address the use objection. Broadly speaking, Smilansky argues that free will skepticism has pragmatic weaknesses that make it difficult to safeguard deontological constraints in a way that our beliefs in free will, moral responsibility, and basic desert can. The incapacitation model is still vulnerable to utilitarian and deterrence theory objections, and it is not enough to merely state that the view does not allow using people as a means. The deontological constraint that it is wrong to use people as a means is one that people believe *because* we believe in desert. Under free will skepticism, basic desert is essentially ruled out. Smilansky argues, "once the difference between offenders and nonoffenders is not held to be reflective of moral guilt and innocence, focusing on it loses much of its moral motivation" (Smilansky, 34). Smilansky's argument highlights the fact that we genuinely believe that it is wrong to use people as a means because we think that people *deserve* better.

Our belief in free will and desert helps preserve justice and is the reason that we believe that using people as a means is wrong. Pereboom argues that his view does better than utilitarian and deterrence theories, because punishing the innocent is unacceptable in his view. However, because free will skepticism destroys the distinction between the morally guilty and the morally innocent, no one is ever technically deserving of anything. Because of this, optimistic skeptics are in a weaker position to "resist practices that are socially efficient but statistically increase the

chances that the innocent will be wrongly prosecuted or convicted" (38). Although Pereboom initially believes his view avoids the problem of punishing the innocent, he still faces this objection because there are no "innocent" or "guilty" people under the free will skeptical view.

Expanding on this objection that Smilansky proposes, I believe that there are more insidious concerns with Pereboom's quarantine and Caruso's public health models. Not only is it the case that under the quarantine model we would be statistically more likely to increase the chances that the innocent will be wrongly prosecuted (because of reasons that Smilansky explains above), but it is also more likely that we would see people who already experience systemic injustices become more susceptible to harm. First, if we were to accept and implement the Public Health Quarantine, we could see people being put into isolation before they have even committed crimes, based on the perceived "threat" that they pose. Invoking the right to self-defense has led people to take actions to neutralize threats before they arise, and based on empirical evidence, we might expect to see the same issue arise under the quarantine model, especially given the fact that Caruso emphasizes prevention.

Plenty of historical examples capture this worry; for example, take the internment of Japanese Americans during World War II. Innocent Japanese Americans were perceived as threats simply due to their ethnicity, and therefore, President Franklin D. Roosevelt ordered their involuntary and inhumane imprisonment. Another recent example of such an injustice is the immoral scrutiny and profiling of Arab Americans, specifically concerning Muslims. After 9/11, a sizable minority in the U.S. wanted the government to target and increase surveillance on Muslim Americans.[10] This was because these people in the public perceived them to pose a

---

[10] Panagopoulos, C. (2006). Trends: Arab and Muslim Americans and Islam in the Aftermath of 9/11. *The Public Opinion Quarterly*, *70*(4), 608–624. http://www.jstor.org/stable/4124213

threat, labeling many innocent Arabs as potential terrorists. Preventative measures can lead to great injustices, as we have historically seen against marginalized groups.

Facts about Caruso's public health model exacerbate this issue. His solution centers the responsibility to address the issues that are determinants of criminality, including poverty, racism, and other systemic injustices. Pairing Pereboom and Caruso's solutions together, we can reasonably worry that there is nothing about the optimistic skeptical view stopping us from looking at the social determinants of criminal behavior and using those indicators to decide who should be quarantined. Given statistics and facts about those who commit crimes, the quarantine and public health model, particularly when coupled together, would most likely target people of color. Particularly those people of color who are poor and or mentally ill.

Consider fearmongering and sensationalist media rhetoric which paved the way for the dramatic increase in mass incarceration in the 1960s. From the progressive era onward, police officers became far more visible on the streets of low-income neighborhoods where the majority of residents were black. Juvenile justice practitioners in northern cities effectively channeled black youth into newly emerging institutions, and we can still see the disproportionate number of black children in the juvenile system today. In 2015, black children made up forty-four percent of detained youth even though they only comprise 16 percent of all youth in the U.S.[11] Black Americans are incarcerated in state prisons at a rate that is 5.1 times the imprisonment of white Americans, and Latinxs are imprisoned at a rate that is 1.34 times the rate of whites (Nellis 2016).

---

[11] M., Sladky, T.J., Kang, W., & Puzzanchera, C. (2017). "Easy Access to the Census of Juveniles in Residential Placement."

Let us consider one current measure that police in the U.S. use: preventive racial profiling. Racial profiling is a preventative strategy that constitutes "any police-initiated action that relies on the race, ethnicity, or national origin and not merely on the behavior of the individual" (Risse, 62). Preventative racial profiling uses statistical evidence of who is likely to commit a crime to initiate police stops and searches to prevent crime. This leads police to stop and search black Americans at a far higher rate than whites (Hetey and Eberhardt 2018).

The state already uses Caruso's social determinants of criminality to justify racial profiling, even though we see that it is plainly unjust. Statistics about the criminalization of the mentally ill worsens this issue, as well; about two in five people who are incarcerated have a history with mental illness.[12] Further, seventy percent of youth in the juvenile system have a diagnosable mental health condition.[13] On top of these disturbing statistics about mental health, there are also worrisome implications when we consider economic inequalities are positively related to high rates of violent crimes in American metropolitan areas. Southern cities have higher rates of criminal violence due in large part to the greater economic inequality there (Blau & Blau 1982). Another comprehensive global study investigating the link between income inequality and crime shows that violent crime is jointly determined by the pattern of income distribution and by the rate of change of national income (Fajnzylber & Loayza 2002). This is almost glaringly problematic when we consider one of Pereboom's responses to the use objection, which is that we should only incapacitate those who commit the most dangerous or violent crimes.

---

[12] Bronson, J. & Berzofsky, M. (2017) Bureau of Justice Statistics, Retrieved 5 May, 2021 from https://www.bjs.gov/content/pub/pdf/imhprpji1112
[13] Cocozza, J., & Skowyra, K. (2007). Blueprint for change: A comprehensive model for the identification and treatment of youth with mental health needs in contact with the juvenile justice system. Delmar, NY: The National Center for Mental Health and Juvenile Justice

Given what we know about people who are more likely to encounter the criminal justice system, and moreover, have a higher propensity to commit "violent" crimes, implementing Pereboom's incapacitation and Caruso's public health-quarantine model would be disadvantageous for black and Latinx people, poor people, and people with mental health issues. It would be especially harmful for the poor, mentally unwell black or Latinx person. Further, given that these views disregard basic desert, it fails to safeguard against deontological constraints that would prevent these people from being detained unjustly or before they commit crimes.

While Caruso and Pereboom assume that their forward-looking models are supposed to be fairer and more just, the disappearance of deontological constraints, foundation in the right to self-defense, and the worries with the use objection and social determinants of criminality show that these models may incur further harms to those who already suffer from systemic injustice. Pereboom and Caruso's endorsement of the *principle of least infringement* here does not help; I am not concerned about the conditions of isolation that the victims of the quarantine model may be subjected to, but I am rather pointing out the injustice that comes with singling them out and determining that they should be quarantined in the first place.

Concluding Thoughts and Looking Forward

In the first section of this essay, I surveyed the main views in the contemporary free will debate. In Section 2, I considered how Pereboom persuasively motivates hard incompatibilism. In Section 3, I discussed Strawson's reactive attitudes and why they might be threatened under free will skepticism. In Section 4, I examined theories of criminal punishment that free will skeptics can appeal to, given that they do not appeal to basic desert. In Section 5, I focused on

Pereboom's quarantine model and Caruso's public health model. Finally, in Section 6, I offered my own objections to Pereboom's and Caruso's recommendations. First, I argued against Pereboom's analogues for certain reactive attitudes by pointing out that these analogues are not as harmless as they seem, and also that they are inappropriate substitutes in many situations. I then highlighted several moral concerns for the quarantine models. Here I argued that features of Caruso's public health-quarantine model can target and cause more harm to already marginalized groups. These concerns urge us to consider whether Pereboom and Caruso's view is as humane as many free will skeptics readily assume.

We have substantial reasons to accept free will skepticism theoretically, and this important consideration means that we should adopt a revisionist view of free will moving forward. However, practical challenges should encourage us to reject Pereboom's analogues to the reactive attitudes. Moreover, Pereboom's and Caruso's solutions to dealing with criminality have problematic implications, especially when taken together. We can, and I believe should, reject retributivism while also rejecting Pereboom's quarantine model based on the use objection. Caruso's argument tracks the social determinants of crime that our justice system fails to sufficiently acknowledge. Morality, justice, and fairness requires that we care about addressing the systemic issues that cause criminal behavior. However, we cannot do this in a just and fair way without preserving our notions of free will, moral responsibility, and basic desert.

Works Cited

Mele, A. R. Free Will and Luck. Oxford: Oxford University Press, 2008.

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

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

Pereboom, D. "Chapter 4: Problems for Compatibilism." Essay. In *Living without Free Will*, 89–126. Cambridge: Cambridge Univ. Press, 2006.

Kane, R. (2011). Chapter 1: The Contours of the Contemporary Free Will Debate. In *The Oxford Handbook of Free Will* (pp. 3–37). Essay, Oxford University Press.

Van Inwagen, P. (1975). The Incompatibility of Free Will and Determinism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 27*(3), 185-199. Retrieved May 3, 2021, from http://www.jstor.org/stable/4318929

Kim, J. (1998). *Mind in a physical world: An essay on the mind–body problem and mental causation*. MIT Press.

Shaw, E., Pereboom, D., & Caruso, G. D. (2019). Chapter 1: Free Will Skepticism in Law and Society: An Overview. In *Free Will Skepticism in Law and Society: Challenging Retributive Justice* (pp. 1–26). Essay, Cambridge University Press.

Morris, H. (1981). Reflections on feeling guilty. *Philosophical Studies* 40 (2):187 - 193.

McCloskey, H. J. (1965). A non-utilitarian approach to punishment. *Inquiry: An Interdisciplinary Journal of Philosophy* 8 (1-4):249 – 263.

Farrell, D. (1985). The Justification of General Deterrence. *The Philosophical Review, 94*(3), 367-394. doi:10.2307/2185005

Pereboom, D. (2013). Free Will Skepticism and Criminal Punishment. In Thomas A. Nadelhoffer

    (ed.), *The Future of Punishment*. Oup Usa. pp. 49.

Powers, M., & Faden, R. R. (2006). *Social Justice: The Moral Foundations of Public Health and*

    Health Policy. Oxford Univ. Press.

Caruso, G. D. (2017). The Public Health-Quarantine Model. *SSRN Electronic Journal*.

    https://doi.org/10.2139/ssrn.3068021

Caruso, G. D. (2016). Free Will Skepticism and Criminal Behavior. *Southwest Philosophy*

    Review, *32*(1), 25–48. https://doi.org/10.5840/swphilreview20163214

Nichols, S. (2007). After incompatibilism: a naturalistic defense of the reactive attitudes*.

    Philosophical Perspectives, 21(1), 405-428. doi:10.1111/j.1520-8583.2007.00131.x

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46*(4),

    352–367. https://doi.org/10.1037/0003-066x.46.4.352

Lazarus R. S. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press

    Nelkin, D. K (2011). *Making Sense of Freedom and Responsibility*. Oxford University

    Press.

Cichy, K. E., Lefkowitz, E. S., Davis, E. M., & Fingerman, K. L. (2013). "You are such a

    disappointment!": negative emotions and parents' perceptions of adult children's lack of

    success. *The journals of gerontology. Series B, Psychological sciences and social*

    sciences, *68*(6), 893–901. https://doi.org/10.1093/geronb/gbt053

Yang, J., & Zheng, Y. (2019). Links Between Perceptions of Successes, Problems and Health

    Outcomes Among Adult Chinese Children: The Mediating Role of Perceptions of

Parents' Feelings and Intergenerational Relationships. *Frontiers in psychology*, *10*, 2551. https://doi.org/10.3389/fpsyg.2019.02551

Watson, G. 1993. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In Perspectives on Moral Responsibility, ed. John Fischer and Mark Ravizza, 119–48. Ithaca, NY: Cornell University Press (orig. pub. 1987).

Shabo, S. (2011). Where Love and Resentment Meet: Strawson's Intrapersonal Defense of Compatibilism. *Philosophical Review*, *121*(1), 95–124. https://doi.org/10.1215/00318108-1426373

M., Sladky, T.J., Kang, W., & Puzzanchera, C. (2017). "Easy Access to the Census of Juveniles in Residential Placement."

Nellis, A. (2016). The color of justice: Racial and ethnic disparity in state prisons. The Sentencing Project. Retrieved from https://www.sentencingproject.org/wp-content/uploads/2016/06/The-Color-of-Justice-Racial-and-Ethnic-Disparity-in-State-Prisons.pdf

Hetey, R. C., & Eberhardt, J. L. (2018). The Numbers Don't Speak for Themselves: Racial Disparities and the Persistence of Inequality in the Criminal Justice System. *Current Directions in Psychological Science*, *27*(3), 183–187. https://doi.org/10.1177/0963721418763931

Cocozza, J., & Skowyra, K. (2007). Blueprint for change: A comprehensive model for the identification and treatment of youth with mental health needs in contact with the juvenile justice system. Delmar, NY: The National Center for Mental Health and Juvenile Justice

Alicia Bannon et al., Brennan Ctr. for Justice, *Criminal Justice Debt: A Barrier to Reentry* 7
(2010)

Fajnzylber, P., Lederman, D., & Loayza, N. (2002). Inequality and Violent Crime. *The Journal of Law & Economics, 45*(1), 1-39. doi:10.1086/338347

Caruso, Gregg, "Skepticism About Moral Responsibility", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL =
<https://plato.stanford.edu/archives/sum2021/entries/skepticism-moral-responsibility/>.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, *68*(1), 5–20. https://doi.org/10.2307/2024717

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*.

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

O'Connor, T. (1995). Agent causation. In *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. Oxford University Press. pp. 61-79.

Hieronymi, P. (2001). Articulating an Uncompromising Forgiveness. *Philosophy and Phenomenological Research*, *62*(3), 529–555. https://doi.org/10.2307/2653535

Chappelle, D. & Angelou, M. Sundance Channel. (2006, November 30). *Iconclasts*. Season 2. Episode 6.

Panagopoulos, C. (2006). Trends: Arab and Muslim Americans and Islam in the Aftermath of 9/11. *The Public Opinion Quarterly*, *70*(4), 608–624. http://www.jstor.org/stable/4124213