In Defense
of
Moral Responsibility Skepticism

by

Jody Tomchishen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Philosophy

Waterloo, Ontario, Canada, 2015

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Moral responsibility skeptics have often focused on problems involving determinism in order to defend their position. I argue that this defense of moral responsibility skepticism is misplaced given that what really matters for moral responsibility is an agent's ability to have morally-relevant control. An account, I call agnostic control, remains viable regardless of the truth of determinism, which means that determinism is the wrong place to look for the denial of moral responsibility. I provide an argument in favour of moral responsibility skepticism, such that differences in ability between agents, which are the result of constitutive luck, are moral responsibility undermining. This is explained by the contrastive fact that agents have differing abilities, which makes praising one agent, and blaming another, inherently unfair. I then defend this skeptical position against three prominent criticisms, that the system of moral responsibility is 'fair enough', that without moral responsibility our participant reactive attitudes would not be justified, and that moral responsibility is required to justify criminal punishment. Contrary to the claim that moral responsibility is 'fair enough' I argue that the system of rewards and punishments, which are justified by moral responsibility, require that moral responsibility have a higher standard of fairness. This is due to the increased significance these rewards and punishments have for the agent. The concern surrounding the reactive attitudes and criminal punishment involves the idea that both are practically necessary for a well-functioning society. The worry is that without moral responsibility the reactive attitudes and criminal punishment would have to be rejected, an outcome which is purportedly undesirable. I address this problem in two ways; either by showing that the purported benefits are not actually beneficial or by showing that the beneficial aspects of each can and should be maintained given moral responsibility skepticism. This means that moral responsibility skepticism remains a viable option.

Acknowledgements

I would like to thank and acknowledge Mathieu Doucet for support and advice while working on this thesis. I would also like to thank and acknowledge the department of philosophy at the University of Waterloo for being welcoming and encouraging. Finally, I'd like to thank my wife Rachel, whose love and patience made this possible.

Dedication

For Darwin Michael Tomchishen.

Table of Contents

Preface

This thesis articulates an argument against moral responsibility. Moral responsibility, in a basic sense, is the view that agents deserve praise or blame, reward or punishment, by virtue of the individual's actions either increasing or decreasing their deservedness. Moral responsibility is often taken for granted since it is deeply involved in an agent's everyday social interactions. People hold individual agents accountable for their actions and dole out rewards and punishments as necessary. The position I will be presenting here is that this system, though widely adopted, is inherently unfair, and that this unfairness should fundamentally undermine the practice of holding people morally responsible.

The first chapter will address the historical debate surrounding determinism and moral responsibility. Determinism has been historically used to argue for moral responsibility skepticism. My view is that this position fails because it ignores the very real extent that agents have morally-relevant control over their actions. Furthermore, there are several options available for those who wish to ground moral responsibility; options that don't depend on the truth of determinism. I present my own option, which I find to be the most compelling, called agnostic control. Agnostic control is the position that morally-relevant control is necessary for moral responsibility; although, as I will explain in chapter 2, agnostic control is not sufficient for moral responsibility. If a theorist's goal is moral responsibility skepticism, the debate about determinism is misplaced, and not convincing, given that morally-relevant control is not undermined by determinism. By providing my own account, agnostic control, I hope to focus the debate surrounding moral responsibility in the direction of where it is most relevant; thus, making it clear where the moral responsibility skeptic should direct their criticism.

The second chapter is a more direct argument against moral responsibility. This argument involves moral responsibility undermining luck, which factors into an individual's constitution. This luck is moral responsibility undermining since it renders the act of praise and blame unfair when comparing the constitutions of differing agents. If one agent is better capable of responding to moral reasons than another agent, and the less capable agent has no way of making themselves as capable as the better capable agent, then it is unfair to treat them equally when it comes to the distribution of rewards and punishments. This means that control is not as directly important for moral responsibility as is the contrastive fact regarding the differing abilities between agents in the morally-relevant reference group. I will then defend this position against the view that moral responsibility, although ultimately unfair, is 'fair enough'.

The third chapter addresses practical concerns regarding the abolishment of moral responsibility. If I am right, and moral responsibility is not justified, some theorists argue that it is best to maintain moral responsibility anyway since it is practically necessary even if it isn't ultimately justified. In other words, even if moral responsibility is not metaphysically justified, it is pragmatically justified given the benefits moral responsibility confers upon our interpersonal relationships and society. There are two ways of addressing these kinds of arguments; the first involves arguing that the desired practices are in fact not practically necessary, and the second involves arguing that even if the practices are practically necessary they can be maintained while denying moral responsibility. The two practices which are often defended in terms of their practical necessity are an agent’s participant reactive attitudes and the institution of criminal punishment. Regarding reactive attitudes I will argue that various aspects of resentment should be abandoned and that they are indeed not practically necessary. However, some attitudes such as forgiveness and gratitude should and can be maintained even if moral responsibility is denied. Similarly, with various aspects of our criminal justice system, I will maintain that it is not clear that moral responsibility is necessary for these practices. Although, if moral responsibility is in fact necessary for these practices, I will further argue that there are more practical, non-punitive, alternatives, thus making criminal punishment unnecessary.

Chapter 1
Agnostic Control

In order to ground moral responsibility, philosophers have appealed to the ability of agents to have control over various aspects of their morally-relevant decision-making processes. The debate about control has often focused on the nature of the universe and the extent to which determinism or indeterminism allows or constrains an agent’s ability to control their actions. What matters, for the context of this debate, is the extent to which these cosmic factors undermine the kind of control necessary for moral responsibility. In order for an agent’s control to be the kind necessary for moral responsibility, the agent must have control over the morally-relevant features which factor into their decision-making. In other words, the agent must be capable of deliberating over the morally-relevant aspects of a particular situation that presents itself; otherwise, there would be nothing to distinguish, for example, a tornado from a person who destroys several homes. Most human beings, unlike tornadoes, can consider the moral worth and consequences of their actions and thus maintain a degree of control over their morally-relevant decision-making processes. The process/ability must also be 'theirs' in the sense that it is not directly controlled by anything else. When an individual performs an action that is in some sense coerced, then it is not clear that their decision was under their control in the sense relevant for moral responsibility. Disagreement over this will involve the extent to which historical factors diminish direct and ultimate control, and whether or not that undermines moral responsibility. Of particular interest here is the extent to which determinism or indeterminism undermines this kind of control.

Relevant to this discussion is the problem of free will. Free will is often conceived as the ability of agents to act such that they can be held morally responsible. This ability relies heavily on the concept of control. After all, it seems unfair, and possibly incoherent, to hold individuals morally responsible for actions they had no control over. Morally-relevant control is necessary for moral responsibility, but it may be the case that not all forms of morally-relevant control are sufficient for moral responsibility. The purpose of this chapter is to assess whether a coherent morally-relevant control is possible given determinism, indeterminism, or both. It is a further question whether this type of coherently available control is the type of freedom-relevant control required for free will and moral responsibility, something which I will address in the next chapter. However, given the significance of this historical debate, and the fact that moral responsibility skeptics remain focused on issues involving determinism (Pereboom, 2001; Harris, 2012), I think it is important to address why these arguments are insufficient. In this chapter, I will argue that there is a form of morally-relevant control available regardless of whether or not determinism is true. If theorists want to ground moral responsibility, particularly in the abilities agents actually have, this would be a good place for them to start. It also means that those looking to deny moral responsibility have to move the debate away from problems surrounding determinism and towards problems involving psychologically realistic accounts of control.

1.1 Brief Overview of the Available Accounts of Moral Responsibility

There are several common positions adopted by theorists that should be highlighted. The first divide relevant for this discussion is between compatibilists and incompatibilists. Incompatibilists find moral responsibility incompatible with a deterministic universe. The motivating intuition held by most incompatibilists is that if the universe is determined, such that an agent's actions are caused by features outside of their control, they would lack the control relevant for moral responsibility. Incompatibilists could be further separated into two groups. The first group, hard determinists, find it compelling that the universe is deterministic and thus they reject moral responsibility (Harris, 2012). The second group, libertarians, are those who find it compelling that there are moments of morally-relevant control given by the indeterministic features of our universe. The three most prominent libertarian positions disagree about the nature of these indeterministic features and thus disagree about the types of morally-relevant control agents possess. The three accounts either view the morally-relevant indeterministic features to be noncausal (McCann, 2012), agent-causal (Clarke, 2003; Pereboom, 2001) or event-causal (Kane, 2009; Mele, 2006).

Not all moral responsibility skeptics are hard determinists. Hard determinists agree with libertarians that if moral responsibility were to exist, then it would require a morally-relevant indeterministic universe. There is another view that can reasonably be described as incompatibilist, but does not endorse this conditional. This view, which is argued for by Derk Pereboom, is called hard incompatibilism. As Pereboom (2001) points out, this position is sympathetic to hard determinism, but has one significant disagreement, such that moral responsibility is also incompatible with the types of indeterministic processes allowed by the current understanding of quantum mechanics (p. xviii). This means, that even though the universe has some indeterministic processes, the indeterminism available does not allow for the type of control proposed by Pereboom's own libertarian account. This distinguishes his view from other accounts, which are described as impossibilist, since his view accepts that some indeterministic features would allow for moral responsibility, but that they are empirically not likely (Pereboom, 2014, p. 3-4). An example of an impossibilist is someone who argues that regardless of the truth or falsity of determinism, or the type of indeterminism available, there is no moral responsibility. This kind of position is most prominently defended by Galen Strawson (1994).

Compatibilists, as the name would imply, find moral responsibility to be compatible with a deterministic universe (Frankfurt, 1969; Dennett, 1984). They argue that determinism allows for, rather than hinders, an agent's ability to control their morally-relevant decision-making processes. Some disagreements between compatibilists involve the extent to which historical features limit an agent's ability to control their morally-relevant decision-making processes, such that they render what control an agent has left insufficient for moral responsibility. Those who find these historical, responsibility undermining features, compelling, for example, the introduction of neuro-manipulation into an agent's causal history, argue in favour of a history-sensitive compatibilism (Mele, 2006; Fischer & Ravizza, 1998). This means that their ability to control is sensitive to the possible addition of nefarious features in their history, such as manipulating neuroscientists, which remove them from being candidates for moral responsibility. History-insensitive compatibilists argue that nefarious features, such as manipulating neuroscientists, should not undermine an agent's candidacy for moral responsibility so long as the manipulation itself does not render us incapable of controlling the morally-relevant decision-making processes which factor into an agent's actions (Dennett, 2003). All that matters for the history-insensitive compatibilist is that the agent remains the type of being capable of morally-relevant control, regardless of the historical processes which factor into the agent's creation.

There also exist agnostic accounts, which maintain moral responsibility regardless of the truth of determinism. The two most common agnostic accounts are semicompatibilism (Fischer & Ravizza, 1998), and agnostic autonomism (Mele, 1995; Mele, 2006). They focus on the aspects of agents which help account for control, without worrying about the ultimate nature of the universe. In this sense they share a lot in common with compatibilists, but agnostics leave open the possibility of an indeterministic account of moral responsibility.

There are still some philosophers who would argue for non-natural indeterministic processes that could be used to ground free will and moral responsibility. I will not directly address such positions, mainly because they lack substantial evidence in their favour. It may be the case that some yet unknown, non-natural, power of free will could be sufficient to ground moral responsibility, but given that the power is non-natural, it is not even clear to what extent evidence can be brought to bear on this issue. In this paper I will only be addressing those forms of free will or morally-relevant control which could be argued for on the grounds of methodological naturalism.

In this chapter, I assess the extent to which a coherent conception of morally-relevant control could be reasonably maintained given the truth or falsity of determinism. In section 2, I will focus on the various incompatibilist views, arguing that there exists a coherent indeterministic account of morally-relevant control called daring soft libertarianism. In section 3, I will focus on compatibilism, arguing that indeterminism is not necessary for a coherent account of morally-relevant control. More specifically, I will argue that this compatibilist account must be history-insensitive. In section 4, I will address views such as semicompatibilism and agnostic autonomism, arguing in favour of my own account, agnostic control. Agnostic control is a history-insensitive account that is agnostic about determinism, and is the most coherent account of morally-relevant control available. If it is indeed worthwhile to ground moral responsibility, then agnostic control is a good place to locate its foundation. Agnostic control is particularly apt as a foundation for moral responsibility considering the current, open, scientific debate about the deterministic status of the universe. Although in this chapter I will argue that agnostic control is necessary for moral responsibility, in the next chapter I will argue that it is not sufficient.

1.2 Incompatibilism

There are two main intuitions used to argue in favour of incompatibilism. These arguments are made based on problems that the truth of determinism would create for what incompatibilists want an agent's morally-relevant control to be able to accomplish. The first problem involves the idea that for agents to have morally-relevant control there must be actual alternative possibilities open to them. Actual, in the sense that there was a real possibility that the agent could have done something differently than what they did. The second problem involves the idea that moral responsibility requires the agent to be the source of their action, something that is claimed to be impossible if an agent's actions are merely the product of a causal chain that was outside of their control (Pereboom, 2001; Clarke, 2003). These two positions have related aspects which involve control. In order for an agent to be free, incompatibilists want that agent to be able to be in control of future possibilities, while at the same time being suitably removed (free) from the deterministic processes that are not under their direct control. It is a separate question whether or not the control wanted by incompatibilists is sufficient to grant the agent moral responsibility, but it is possible to assess whether an incompatibilist agent having control, by satisfying these intuitions, is coherent in the first place.

Therefore, a coherent incompatibilist view must account for these two criteria; leeway (LW) and source (S) (adapted from Pereboom, 2001):

> (LW) An action is under the agents control only if the agent could have done otherwise than they actually did.
>
> (S) An action is under the agents control only if it is not produced by a deterministic process that traces back to causal factors beyond the agent's control.

It is possible that these two criteria are not necessary for moral responsibility, but for those who find them intuitively compelling, an account which ignores them will be unsatisfactory. I will now address whether a coherent account is available that satisfies these criteria.

1.2.1 Noncausal Incompatibilism

Three theories have been proposed to explain how incompatibilist control can be accounted for by satisfying both LW and S. The first of those theories involves a type of agency that is noncausal. This view attempts to differentiate acts of will from standard cause and effect. Hugh McCann (2012) explains this difference by contrasting the intrusiveness of an agent's will with the smooth procession of causal chains:

> The object ball's acceleration in the game of billiards is in no way an intrusion: it is an outgrowth of the past whose nature can be understood if only we understand the nature of the conditions preceding it, and whose occurrence can fully be anticipated based on these conditions if only we are prepared to assume that the world will continue to be. My decision to take the rural route to the airport is nothing like this, even if we allow for whatever differences we may think pertain to the mental realm as compared with the physical. My act of deciding *is* an intrusion. (p. 258)

This intrusion, into the causal chain, by an active agent which is not itself a part of any causal chain, is the focus of noncausal accounts. What proponents of noncausal accounts of moral responsibility must explain is how their theory accounts for the control problem that worries other incompatibilists. By severing aspects of the causal connections involved in an agent's deliberative processes, in what sense can control be maintained?

Proponents of the noncausal theory must provide an explanation for how intentional decisions manifest in an agent's behaviour if they are noncausal. As Alfred Mele (2009) explains, agents want their intentions to play a causal role in the production of their actions or else their intentions would not be effective. It is almost incoherent to think that there can be noncausal intentions, for if the agent is in control there must be some causal description between their bodily movements and the intentions that were formed (Mele, 1995, p. 10-11). There would also be a problem for noncausal proponents if they were to argue that, although intentions cause bodily movements, intentions themselves are noncausal. After all, agents want their deliberations to contain reasons which help to cause their actions, and if these reasons to act did not play a role in the formation of their intentions, in what sense can it be said that agents control their actions? Without an answer to this question non-causal accounts remain incoherent.

1.2.2 Agent-Causal and Event-Causal Incompatibilism

The other two theories used to account for incompatibilist control are related to each other in that they accept causation into their overall framework. The agent-causal view is more closely related to the noncausal view, in that the agent itself is not caused by preceding events. However, whereas the noncausal view fails at explaining how willing becomes manifest in action, the agent-causal view accepts the agent as a cause for future actions. Event-causal proponents accept that events play a causal role in an agent's decision-making processes, since in order to appropriately have control over something their reasons must somehow factor into their decisions. The event-causal proponents differ from compatibilists, in that they still feel that at the very least a small amount of indeterminism in the universe will satisfy the LW criterion. The conflict surrounding agent-causal and event-causal accounts of control is the extent to which the added ability in the agent-causal account provides more robust reasons for thinking that this type of agent has more control over their history. The debate, then, is about which view more realistically accounts for S.

Robert Kane (2009) defends a version of the event-causal account of moral responsibility. In his view, humans must be the ultimate originators for at least some of their decisions or actions in order to be held morally responsible (p. 35). This power, of being the ultimate originator of their actions, allows for ultimate responsibility (UR) by accounting for instances of self-forming actions (SFAs), which are not deterministic. In other words, UR, and by extension SFAs, account for moral responsibility, since the agent is the sole source of their decision-making. This is because the agent becomes, in these moments, the ultimate cause of their actions. The agent is free from any deterministic causal chains. SFAs occur at moments when agents are torn between two competing decisions, both of which they have reason to perform. These internal chaotic moments are then satisfied by indeterminate micro-physical processes at a level below our neuronal processing (p. 39). Kane's view allows for there to be other forms of freedom that are more compatible with determinism, such as the influence of reasons on an agent’s actions, but that only the indeterminism provided by SFAs has the type of freedom required for UR. This is because SFAs satisfy LW and S.

Pereboom (2001) argues that Kane's SFAs are not robust enough to give an adequate account of S. In particular, he thinks that S is more fundamental than LW. Although event-causal views have some connection to S they are not as robust as agent-causal accounts in this respect (p. 33). This is because S has a related intuition that appeals to agent-causalists, which he refers to as origination (O) (p. 4):

> (O) If an agent is morally responsible for their deciding to perform an action, then the production of this decision must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which they have no control.

O stands as a challenge to event-causal theorists because it offers an enhanced conception of control, without which, they argue, event-causal incompatibilism is incoherent. Kane's (1999) view is that all the standard causal aspects of history play their role in influencing an agent’s SFAs, but that at the moment in which the SFA is occurring the agent’s effort and the indeterminism are infused. The indeterminism could not occur before or after the effort, or it wouldn't be the agent’s (p. 232). The problem with this move is that it seems like the indeterminism is being applied arbitrarily, since most of the work attributed to the agent has little bearing on the final outcome of the indeterministic process itself. In other words, the kind of LW on offer, by event-causal theorists such as Kane, is not produced by a source over which the agent has control, and thus does not satisfy O (Pereboom, 2001, p. 52-53). Kane (2009) accepts this conclusion by admitting that the indeterminism that allows for possibility can also act as a hindrance when it comes to control (p. 44). However, control seems to be a robust intuition, in part shared by both compatibilists and incompatibilists, although they may disagree about how much control is necessary, or what an agent must have control over to be morally responsible. Removing control altogether, especially control over something so integral to an agent’s character development, runs counter to those intuitions. This concern is known as the disappearing agent problem, and is a problem that weakens Kane's event-causal view.

A proposed solution to this problem is to argue for an agent-cause. An agent, as a special substance, can be situated outside of an event-causal history and can therefore be in the position to make decisions, which are not caused by preceding events. Randolph Clarke (2005) argues in favour of this type of agent, which he describes as an uncaused cause (p. 409). For this to succeed, an agent-causal account must consider to what extent the agent-cause is indeterministic and in what way this indeterminacy distinguishes it from event-causal incompatibilism. Clarke is not forthcoming about what this agent substance is and how it can differentiate itself from event-causation, but offers a concise description of what must be explained to account for it

> If a substance has a certain causal power in virtue of possessing a property that stands in a certain relation to other properties, how can the exercise of that causal power be anything other than *the substance's possessing that property's* (an event's) standing in the causal relations to an effect, as it is in cases of the manifestation of causal powers carried by other properties? (2003, p. 192)

Pereboom (2014) believes he has an answer to Clarke's question, by invoking the possibility of an agent that resembles the stoic conception of a hegemonikon (p. 57). A hegemonikon is an agent with the executive power to determine which states (or events) will result in action. This agent is also independent of these states (or events), and therefore is not these states (or events). Pereboom then argues that this position is negatively conceivable, such that it cannot be ruled out a priori (p. 58).

Part of the problem inherent in advocating for extra powers is that the powers don't really explain what is required of them to explain. As Mele (2006) puts it "One can enhance a collection of powers that is not up to the task of securing a capacity for free and morally responsible action and get an enhanced collection that also is not up to that task" (p. 68). In what sense does a hegemonikon like agent have control over the indeterminism which factors into their choices and actions? This is particularly confusing if they are not influenced by the events which should reasonably factor into their decision-making. There seems to be an unavoidable infinite regress in trying to place the agent above the reasons and events which factor into their decision-making. In order to be in control of the indeterministic processes which factor into the executive power, an agent would need executive power*. Executive power* determines the states, which determine the states which result in action. But this still leaves unexplained what controls the indeterministic processes which factor into executive power*. An agent who is independent of the states (or events) which factor into their deliberation is still required to be the type of agent who is not determined, and thus it remains open as to why *this* indeterminism does not amount to the same problem found in Kane's event-causal view. The problem of the disappearing agent remains. An incompatibilist, to have a coherent account of control, must account for the agent having some influence over the indeterministic processes which result in action, since these indeterministic processes are morally-relevant.

Indeterministic processes are fundamentally something that agents lack ultimate control over, since to have ultimate control would render the process deterministic. In order to address this problem an incompatibilist must find a way to add at least some control over these indeterministic processes. One way that an incompatibilist could have some control over their decisions and actions is if their cognitive abilities could somehow impact the probabilities of the outcomes of the indeterministic processes. The problem with agent-causation is that if the agent is independent of the events which factor into their decision-making, like Pereboom's hegemonikon, then it is unclear what is motivating the causal power; there is nothing to influence the probabilities of future outcomes. The content of an agent's deliberations inhabit an intimate role with their decision-making and any attempt to sever this relationship is beginning to look a lot like noncausalism. Noncausalism, as I argued in the previous section, is incoherent, so any move in that direction is problematic. By requiring a special type of independence, the agent-causal theorist is setting impossible demands on control. In order to have self-control an agent's reasons (events) must be causal, otherwise, there is no coherent way to account for the disappearing agent problem. Agent-causation fails to be a priori possible, since it fails to coherently explain how their proposed ability controls or impacts the probabilities of the indeterministic processes, which result in action.

This means that it is possible to maintain that there are event-causal accounts, which willingly incorporate reason informing control. One such view is Mele's (2006) daring soft libertarianism (p. 111-117). This event-causal account differs from Kane's, since the indeterminism involved is more directly controlled by the agent. Kane's view is that SFAs occur when an agent is caught between two equally compelling options, which means that nothing the agent does directly determines the outcome of this indecision. Daring soft libertarianism has a more robust account of control than Kane's event-causal view, but has the added benefit of being less demanding than an agent-causal account. In order to account for how an agent's reasons impact the probabilities of the outcomes of the indeterministic processes, Mele proposes a mechanism, similar to a roulette wheel, in which the slots corresponding to possible decisions are modified by the agent's reasons. The slots on the wheel are larger or smaller depending on the probability that they will be activated based on previous deliberations and the weights of various reasons. The proximate need to make a decision then releases a ball onto the wheel – or fires up the relevant neuronal equivalent –

which indeterministically chooses which decision will be activated based on where the ball stops. This is an iterative process, and is therefore susceptible to further change and influence through the introduction of different reasons and changing contexts. This iterative process accounts for S, since the deterministic features of the daring soft libertarian account are not wholly determined by events outside of the agent's control.

Some may view daring soft libertarianism as being too close to compatibilism. Although daring soft libertarianism accounts for LW, it lessens the extent of ultimacy found in S by allowing some features of an agent's decision-making processes to be influenced by factors outside of their control. However, if ultimacy is incoherent, demanding ultimacy is tantamount to conceding to the moral responsibility skeptic. If there is a need to ground moral responsibility, demanding something impossible is a bit extreme. As Daniel Dennett (2003) points out

> If you are one of those who think that free will is only *really* free will if it springs from an immaterial soul that hovers happily in your brain, shooting arrows of decision into your motor cortex, then, given what *you* mean by free will, my view is that there is no free will at all. If, on the other hand, you think free will might be morally important without being supernatural, then my view is that free will is indeed real, but just not quite what you probably thought it was. (p. 222)

I do not think free will is morally important, as I will argue in chapter 3, but otherwise I think Dennett is right on this point. If you are inclined to think free will is important, then you better not set the bar too high, such that it would be a priori impossible. Agents will never have control over all things, or even over all aspects, of one dimension, of life, such as their morally-relevant decisions. However, what agents can have is some control over some aspects of their life; in particular, some control over the morally-relevant parts of the agent's decision-making processes. If agents can be held morally responsible, then it is in this range of control, and not ultimate control, which will allow them to be held as such.

Daring soft libertarianism differs from compatibilism by endorsing LW, but more closely approximates compatibilism by lessening the need for ultimacy in S. By satisfying both LW and a less ultimate S daring soft libertarianism is a coherent possibility for an incompatibilist version of morally-relevant control.

1.2.3 Hard Incompatibilism

All of this, of course, does not mean that the hard incompatibilist thesis fails; the thesis that argues that moral responsibility is incompatible with both determinism and indeterminism, since it does not rely on the coherence of agent-causal incompatibilism. Although, if Pereboom were to give up this tentative support for agent-causal theories, it will become harder to distinguish his view from Galen Strawson's impossibilist, 'no free will no matter what', account.

Hard incompatibilism addresses the fact that free will, and thus moral responsibility, do not fit with either determinism or indeterminism. I have sympathy with this view, as will be clear in the next chapter regarding responsibility undermining luck. However, my disagreement with Pereboom involves the extent to which free will, or by extension morally-relevant control, can be ruled out given the truth of determinism. This disagreement is over the persuasiveness of the various incompatibilist or compatibilist intuitions. I find it hard to conclude that someone's intuitions are wrong or right, especially when they are about something as inconclusive as whether moral responsibility requires a specific state of the universe. Using intuitions to argue against intuitions is the equivalent of pounding the table during an intellectual stalemate to score a point; it won't get you very far. This is not to say that progress can't be made, but I don't think discounting the varying intuitions about freedom and morally-relevant control is a good move in this debate. Discounting these intuitions is particularly problematic if there is no conclusive way of ruling out these intuitions altogether.

That being said, some intuitions are far too demanding, and it is possible that Pereboom's intuitions regarding freedom are of this sort. Something like O demands a level of control which many would deem too stringent. The past is not up to an agent because the agent lacks control over it, but as Mele (2006) argues “...it does not follow from (1) the supposition that these things are not up to us in this sense and (2) the supposition that our actions 'are the consequences of' these things, that (3) our actions are 'not up to us' in the *same* sense” (p. 138). Agents still, in some sense, control their present, even if they have no control over their past. Demanding that agents have this kind of control for moral responsibility requires a further argument in support of this demand.

As I have previously argued, daring soft libertarianism is a coherent account of control for incompatibilists by satisfying the two incompatibilist intuitions, LW and a less ultimate S. Kane's event-causal view lacks direct control over the probabilities of the outcomes of the indeterministic processes, and is therefore insufficient for morally-relevant control. Noncausal and agent-causal accounts are both incoherent because they require something impossible in order for moral responsibility. By being too demanding about ultimate control, they are indirectly conceding to the moral responsibility skeptic. This concession would be unsatisfactory for those who advocate for moral responsibility and would not be convincing if offered as the only option by a moral responsibility skeptic. In the next section I will address whether or not compatibilism, by giving up LW and S, could also be maintained as a coherent conception of morally-relevant control.

1.3 Compatibilism

Compatibilism is the position that the truth of determinism does not threaten moral responsibility. A large part of what motivates compatibilism is a rejection of incompatibilist intuitions. If you remember from the last section, these incompatibilist intuitions are:

> (LW) An action is under the agents control only if the agent could have done otherwise than they actually did.
>
> (S) An action is under the agents control only if it is not produced by a deterministic process that traces back to causal factors beyond the agent's control.

One compatibilist strategy is to find examples where an agent could not have done otherwise and yet is still intuitively in control of their actions, therefore finding fault with LW. The second and most comprehensive approach is to discuss ways in which agents have control in deterministic universes, such that they are the source of their decisions, therefore finding fault with S.

1.3.1 Frankfurt-Style Cases

In an influential paper, Harry Frankfurt (1969) argued that there are cases in which agents can be held morally responsible for their actions even though they could not have done otherwise. Frankfurt's original case involves Black, who wants Jones to perform an action, and who has it within his power to ensure that Jones performs it. Now, Jones does perform the action, and as it happens he does so without any intervention by Black. Had Jones given a sign that he was not going to perform the action, however, Black would have intervened to ensure that Jones did in fact perform it. In this case, it seems that Jones is morally responsible for what he did, since he was the source of his decision to act, even though, due to the possibility of Black's interference, Jones could not have done otherwise (p. 835).

Frankfurt thinks moral responsibility is maintained in this case because, in some relevant sense, Jones is still the source of his actions. This is because Jones' beliefs, intentions, and other relevant cognitive processes were the direct reasons for why the action was performed, and that even if it was not technically open to him to do otherwise, Jones is the sole source of the action performed. Since Jones was responsible even though he could not have acted otherwise, Frankfurt concludes that alternate possibilities are not necessary for moral responsibility. Without getting too involved in whether or not this actually grants Jones moral responsibility, it is enough to suggest that what is motivating the worry for incompatibilists is the supposed necessity of the LW intuition. If Frankfurt is right LW is superfluous, and alternate possibilities are not a necessary component of morally-relevant control.

There are two types of incompatibilist rejoinders to these Frankfurt-style cases. The first is to argue that even though something close to LW is true, S is the more important incompatibilist intuition, and it is not clear that S is diminished in these cases. Although the Frankfurt-style cases show LW to be unnecessary for moral responsibility, they do not undermine S, and therefore incompatibilism remains a live option. Pereboom (2001) distinguishes between two types of incompatibilists, those who prefer LW over S, leeway incompatibilists, and those who prefer S over LW, causal history incompatibilists (p. 5-6). In the case of Jones, considering that the internal causal history of his action is still in some sense 'up to him', Jones is still in control in the incompatibilist sense involving the S intuition, even if he could not have done otherwise in the grander scheme of things. This is called the causal history response to Frankfurt-style cases. The other incompatibilist response is to argue that some elusive form of determinism is still present in the Frankfurt-style cases, and, therefore, a leeway incompatibilist may not find the cases intuitively appealing. This can be called the dilemma response to Frankfurt-style cases.

The best argument in favour of the dilemma response to Frankfurt-style cases is offered by David Widerker (1995), which involves the extent to which a manipulator can know what an agent would do based on some prior sign. Suppose the tell, or prior sign, which signals Black not to intervene, is Jones' blushing, and the action was for Jones to kill Smith. Then it will follow that “If Jones is blushing at t1, then, provided no one intervenes, he will decide at t2 to kill Smith” (p. 249). Widerker's dilemma is that if the presence of the prior sign at t1 (the blushing) is causally sufficient for the decision at t2, then the presence of Black does play a role in bringing about t2, but if the blushing at t1 is not causally sufficient then there remains some sense in which Jones could have done otherwise. If t1 is causally sufficient, then it would not be accepted by incompatibilists as an undetermined action, since between t1 and t2 there is no sense in which Jones could have done otherwise. However, if you allow that Jones could have done otherwise between t1 and t2, then the case no longer rules out alternative possibilities and thus would not count as a case against LW. In other words, the introduction of Black into the decision-making process either renders the whole situation deterministic or still allows for LW. If the situation is deterministic, then based on S, incompatibilists will not find the conclusion appealing, but if the situation still allows for LW this renders Black's power inconsequential since the possibilities remain open. Due to this dilemma compatibilists need to construct a case where the blushing or prior sign that signals Black does not render the relevant physical properties, during the time span between the signal and the action, deterministic.

Two solutions have been offered to deal with Widerker's dilemma. The first places the incompatibilist undetermined action before the prior sign. The second is to try to avoid the idea of the prior sign all together. The first of these solutions fails because it simply pushes Widerker's dilemma back in time. If at the time before the prior sign the incompatibilist undetermined action were such that there were alternative possibilities, then the agent could have done otherwise. If you instead reject this conclusion, then the incompatibilist could argue that it wasn't truly an incompatibilist undetermined action and Widerker's dilemma remains (Pereboom, 2001, p. 11-12).

Mele and Robb (1998) offer an elaborate case in which there is no prior sign:

> At *t1*, Black initiates a certain deterministic process *P* in Bob's brain with the intention of thereby causing Bob to decide at *t2* (an hour later, say) to steal Ann's car. The process, which is screened off from Bob's consciousness, will deterministically culminate in Bob's deciding at *t2* to steal Ann's car unless he decides on his own at *t2* to steal it... The process is in no way sensitive to any "sign" of what Bob will decide. As it happens, at *t2* Bob decides on his own to steal the car, on the basis of his own indeterministic deliberation about whether to steal it, and his decision has no deterministic cause. But if he had not just then decided on his own to steal it, *P* would have deterministically issued, at *t2*, in his deciding to steal it. Rest assured that *P* in no way influences the indeterministic decision-making process that actually issues in Bob's decision. (p. 101-102)

This solution does seem to solve Widerker's dilemma, since there are no prior signs which are causally sufficient. However, there may be some push back from leeway incompatibilists who are more inclined towards LW than S. Process *P* does not necessarily interfere with the causal history, but something about it seems to render the causal history "essentially deterministic" (Pereboom, 2001, p. 18).

This would be a problem for causal history incompatibilists since determinism undermines S. To make the “essentially deterministic” feature clear, Pereboom (2001) describes two cases in which an atom follows a causal path with the ability to indeterministically swerve. In both cases the atom does not swerve from the causal path. Although one case has open possibilities, the other has a barrier that would have prevented the swerve, thus eliminating LW. Even though the causal history between both cases is the same, the blockage is a feature which renders the situation outside of the agents control and thus “essentially deterministic” (p. 17). This is only worrying if the leeway wanted requires control over the entire situation. This is too demanding for the obvious reason that agents lack ultimate control. Pereboom's conclusion is that LW is simply entailed by the causal history of agents having certain indeterministic features internal to those agents, and thus causal history incompatibilism is the most coherent way of handling these Frankfurt-style cases (p. 37). By diminishing the type of LW agents have, we end up saving the more important incompatibilist criterion, S.

An incompatibilist who is strongly committed to LW can continue to demand that it is necessary to have these more robust kinds of alternative possibilities, but they need to explain why these robust alternative possibilities are necessary for morally-relevant control. The worry offered by Dennett regarding setting the bar too high applies here as well. If you establish that ultimate control is necessary for moral responsibility at the start of your inquiry, then you have already conceded to the moral responsibility skeptic, given that ultimate control is impossible. If the intuition derived from the Frankfurt-style cases, such that moral responsibility remains while LW is diminished, is correct, then robust alternative possibilities are not necessary for moral responsibility, which, I think, is the correct way to read these cases. This means that Frankfurt-style cases only succeed at undermining leeway incompatibilists, leaving open the possibility for causal history incompatibilism, which is consistent with a daring soft libertarianism.

1.3.2 The Consequence Argument and Manipulation Cases

What remains for the compatibilists is to find some way of undermining the incompatibilist intuition S. The goal is to show that in a deterministic universe control is not undermined or to show that incompatibilism is an impediment to control. The first half of this disjunct is a softer claim, since it would merely show that in some sense the truth of determinism is irrelevant to our ability to self-control, making incompatibilism unnecessary but an open possibility. The second half of the disjunct is stronger; since it would show that indeterminism is not only irrelevant but positively interferes with an agent's ability to control. Daniel Dennett (1984; 2003) offers some support for this side of the disjunct, arguing that determinism allows for what he calls "evitability", whereas indeterminism does not.

Dennett (2003) does not think that anything is gained, in terms of morally-relevant control, through the addition of indeterministic processes. The worry incompatibilists seem to have is that if determinism were true, then everything would be inevitable. Dennett argues that it is the result of the deterministic process of evolution that allows for such things as avoidance. Creatures that were better able to avoid being eaten, or who out smart their prey, are more likely to have offspring, and thus flexibility, in terms of actions and decisions, evolve. This means that determinism does not equate to inevitability, but instead allows for the evolution of what he calls evitability (p. 36-47). When it comes to determined events, agents can see a train coming, understand what will happen if they do not move, and then make a choice about whether it would be better to stay or move. This ability to avoid things, which has been instilled in human beings through the steady process of evolution, allows for opportunity, and thus creates evitability. Dennett puts this in stark contrast with indeterminism:

> It is worth noting... that just as evitability is compatible with determinism, inevitability is compatible with indeterminism. Something is inevitable *for you* if there is nothing *you* can do about it. If an undetermined bolt of lightning strikes you dead, then we can truly say, in retrospect, that there was nothing you could have done about it. You had no advance warning. In fact, if you are faced with the prospect of running across an open

> field in which lightning bolts are going to be a problem, you are much better off if their timing and location are determined by something, since they *may* be predictable by you, and hence avoidable. (p. 60)

Indeterminacy in natural events is not necessarily what incompatibilist have in mind when they are trying to locate freedom enhancing indeterminism, but even if incompatibilists try to move this indeterminism inside the agent it is not clear what this adds. As Dennett (1984) is fond of saying[1] “If you make yourself really small, you can externalize virtually everything” (p. 143). By attempting to try to make the decision 'yours', incompatibilists try to externalize every factor that plays a role in the decision-making process, so that nothing, such as reasons or history, causally determines the agent's actions. The most likely form of an indeterministic event that has some empirical support is quantum indeterminacy, but how is placing a quantum event, internal to the agent, probably in their brain, going to give them morally-relevant control? In other words, how do you make quantum events ‘y*ours’*? This is the dilemma at hand “...if the decision is undetermined – the defining requirement of libertarianism – it isn't determined by you, whatever you are, because it isn't determined by anything” (Dennett, 2003, p. 123). The more an agent begins to incorporate other factors, such as reasons and history, into their deliberative processes the less indeterminate the process becomes. This is what incompatibilists need to account for. They have to have some reason for making the agent undetermined in this way that isn't in some sense arbitrary. They also have to explain how making a causal history undetermined somehow enhances control, rather than hindering it. Pereboom's and Clarke's solution in the previous section was to introduce an agent-causal substance, but as I argued before it is not clear that this kind of special substance is coherent. Furthermore, it is not clear that Dennett’s argument undermines the daring soft libertarian account sketched in the previous section. If you have control over the probabilities of the indeterministic processes which result in action, then in the same compatibilist morally-relevant sense they can be 'yours'.

[1] Dennett first quoted this line in *Elbow Room*, and explained in *Freedom Evolves* that it was the most important sentence of that book (p. 122)

This sort of argument used by Dennett against incompatibilism could be co-opted to argue against compatibilism. If we make everything an agent does determined by the laws of physics, then we remove all control over the particular parts of an agent’s life that are necessary for moral responsibility. If agents are merely the product of natural processes, which began before they were born, which they had no control over, then in what sense can they be in control of how they turned out? This seems to open the possibility that neuro-manipulators and brain washers do not diminish the extent to which agents are morally responsible, since they are simply a part of the causal chain; no different than other social or genetic factors, which influence how individuals develop. This seems to be intuitively problematic, although, as I will argue, not as problematic as some would think. What needs to be addressed is whether compatibilist control is sufficient for moral responsibility, without needing to appeal to an indeterministic process within the agent, in order to stave off the creeping influence of the deep past.

Several views have been presented which attempt to make decisions 'ours' in a determined universe, such that they allow for moral responsibility. Frankfurt (1971) offers a hierarchical account of moral responsibility, which is given by our second-order desires, which are our desires about desires. This is supposed to reveal the agent's authentic choices, reflecting their deep self, since they are the choices the agent truly desires as opposed to desires which are compulsive or the result of coercion. However, there is reason to think this is problematic because there is no reason to suspect that an agent’s second-order desires can’t be just as corruptible as an agent’s first-order desires (Mele, 2006, p. 170-172).

Another view to make decisions 'ours' in a compatibilist sense is to argue for a mechanism within the agent that allows for reasons-responsiveness. This type of view is promoted by both Dennett (2003) and John Fischer and Mark Ravizza (1998). Reasons-responsiveness accounts for the agent’s ability to make decisions based on reasons, which is compatible with some forms of determinism. Agents who have this ability are capable of 'taking responsibility' for their actions by endorsing the actions and decisions that emanate from this capacity. Furthermore, ‘taking responsibility’ or endorsement is something conferred by that same reasons-responsive mechanism.

Pereboom (2001) argues that these various compatibilist positions are not sufficient to explain moral responsibility, since they would fall short of satisfying what he calls the Causal History Principle (CHP) (p. 54). The CHP is as follows:

> (CHP) An action is free in the sense required for moral responsibility only if the decision to perform it is not an alien-deterministic event, nor a truly random event, nor a partially random event.

To motivate this incompatibilist account, Pereboom compares different types of causal histories which maintain both hierarchical authentic deep self compatibilist intuitions as well as reasons-responsive compatibilist intuitions, but changes the degree to which these histories have been manipulated. The goal is to make the worry about the deep past inescapable by these, or any, compatibilist accounts (p. 110-117).

The first case is to highlight an instance of a clearly compromised scenario in which the agent involved intuitively should not be held morally responsible. The goal is to have the universe portrayed in the example be deterministic while maintaining various conceptions of compatibilist control. The case involves a team of neuroscientists that are manipulating an individual named professor Plum by continuously providing him with reasons to kill Ms. White. In this way, Plum remains capable of responding to reasons in the compatibilist sense, which means under that particular conception of control he is morally responsible for killing Ms. White. Plum also maintains a second order desire to kill Ms. White, such that he wants to want to kill Ms. White. The point is that no matter how many compatibilist intuitions are available, it would seem that the intervention by the neuroscientists intuitively render Plum not responsible. Pereboom believes that this intuition hinges on the CHP, since the intervention by the neuroscientists is alien-deterministic.

The second case is identical to the first, except that the neuroscientists manipulate Plum at the time of his creation. That means that they do not have direct control over Plum throughout his life, but their initial manipulations are the factors which causally determine that Plum will kill Ms. White. Again, this seems to intuitively suggest that Plum is not responsible for the death of Ms. White, even though he satisfies various compatibilist intuitions of moral responsibility. This is because his decision to kill Ms. White is entirely within the control of the neuroscientists, who effectively manipulate him to carry out their intentions. The time separating the two different styles of intervention does not seem to be a relevant feature that would make one case responsibility undermining and the other not.

The third and fourth cases are where Pereboom's argument is supposed to undermine compatibilism. In case three, rather than being manipulated by a team of neuroscientists, Plum's behaviour is the result of an invasive cultural upbringing, which has a direct connection to his decision to kill Ms. White. Pereboom argues that in order for this case to be treated differently than the first and second, there must be a relevant feature which explains this difference, a feature Pereboom thinks is lacking. Considering that Plum maintains various components of the compatibilist accounts in all three cases, it is not clear to Pereboom what kind of compatibilist feature would explain the difference in intuitions between the first two cases and the third. If he is correct, then all three cases would be responsibility undermining, which is not a good sign for the possibility of compatibilist morally-relevant control.

The fourth case is simply a more standard case of determinism, such that since the beginning of the expansion of the universe there was no instant where Plum was not going to murder Ms. White. Pereboom's argument is intended to show, or at least motivate the intuition, that there is no morally significant difference between any of the cases that would make moral responsibility appropriate in one case but not in the others. Those who agree that Plum is not responsible in case 1 should, on Pereboom's view, also agree that Plum is not responsible in case 4. Since both cases three and four are alien-deterministic, and also maintain various compatibilist accounts of moral responsibility, it is up to the compatibilist to explain why case four, and possibly three, are not responsibility undermining, while cases one and two intuitively are. More specifically, what is the morally relevant feature that would differentiate the various cases and their differing intuitions? Plum's lacking control over neuroscientists performing the direct manipulation in cases one and two is equivalent to the lack of control he has over the cultural processes which shaped him in case three. The presence of neuroscientific agents, or of culture, or even of a machine that determines Plum's future, are all things which the agent lacks control over and they directly shape how the agent develops. The same sort of control is lacking when considering the universe as a whole, and yet compatibilists seem to think this type of control is not important. It is up to compatibilists, Pereboom insists, to explain why.

There are two responses to this argument. The first is that there is a relevant difference, such that moral responsibility is sensitive to the types of histories which influence an agent’s development. The second is to argue that moral responsibility is maintained in all four cases. This second argument, which I think is the correct response to Pereboom’s manipulation argument, will probably appear as intuitively problematic, but is, I will argue, the only solution to the deep past problem. In order to show why this is the case, I need to both argue that the history-sensitive solutions fail at saving compatibilist moral responsibility, while at the same time providing justification for why Pereboom’s intuitions about the manipulation cases are wrong.

Mele (2006) argues that Pereboom's intuitions, regarding the responsibility undermining effects of each case in his manipulation argument, are misplaced (p. 138-144). If the neuroscientists in cases one and two, in addition to Pereboom’s initial manipulation scenario, instil in Plum the chance that he would be indeterministically incapacitated, then if Plum were to still kill Ms. White the intuition remains that Plum would not be responsible. This means that the source of the intuition is not that cases one and two are the result of deterministic processes. Although I agree with Mele that this is the case, it still does not remove the worry that the past, in a quasi-deterministic event-causal sense, has a diminishing effect on moral responsibility due to its influence on the present. Another argument made by Mele is that the relevant features that distinguish cases one and two from three and four are the introduction of manipulators into the causal history of Plum. This is the basis of Mele’s history-sensitivity; however, Mele's attempted solution for defending compatibilism against manipulation cases concedes too much to the incompatibilist.

Mele’s history-sensitive compatibilism is supposed to take into account the responsibility undermining effects of neuro-manipulators. What differentiates cases of manipulation from normal causal histories, according to Mele, are the processes in which the agents unsheddable values are constructed (p. 164-173). Unsheddable values are not easily lost, and are autonomously constructed through the character building processes of a normal causal history. If values are recently produced which bypass an agent’s control mechanisms, replace their previously autonomous value constructions, and override their ability to consent to the procedure, then the agent is not responsible for the actions these values produce.

According to Dennett (2003), who takes a history-insensitive approach, arguments involving manipulators avoid addressing the differences between neuro-manipulation cases and standard cases of moral education. The problem, it seems, is that there are lots of historical processes which non-autonomously inform and help mould an agent’s development. The only difference between moral education and neuro-manipulation, according to Dennett, is the extent to which the agent is aware of the truth of their causal history. In the manipulated case the agent is not informed about the true history of their production (p. 281-284). Therefore, ‘informedness’ should be a condition for moral responsibility since it allows the agent to either reject or accept their manipulation. This means that ‘taking responsibility’ would require being informed about the true nature of the creation of the agent’s reasons-responsive mechanism. Mele (2006) worries that this means that some neuro-manipulated agents are still responsible for their actions, which he feels “is to bite an extremely hard bullet” (p. 176). Although I find informedness useful in exposing ways in which Mele’s solution fails, it raises additional problems for moral responsibility. For example, most agents are not directly aware of all the causal factors which influenced them during their upbringing, but being ignorant of those facts should not impact the extent to which the agent is held morally responsible. What is useful about the notion of informedness is that if an agent is reasons-responsive, then the information regarding their manipulation can help the agent overcome the procedure. That means, unless the reasons-responsive mechanism itself has been disabled by virtue of the manipulation procedure, the agent can still be held morally responsible.

The main problem with any form of history-sensitivity can be summed up by a Dennett inspired aphorism. If you make yourself too large you will become one with the universe. In other words, there needs to be a cut-off point between being too small, such that everything is external to you and being too large such that you are indistinguishable from everything else. If the agent is too small then there is nothing the agent does that can reasonably be moral responsibility conferring. Similarly, if the agent is too large then it becomes hard to distinguish the agent's actions from the causal histories which influence them. Thus it is hard to distinguish between what the agent produces from what their history produces such that judgements regarding moral responsibility could be reasonably ascribed to the agent. I think the least arbitrary cut-off point which ameliorates this problem is the agent's reasons-responsive mechanism. This is because it maintains a tight connection with the individual's body, as well as being the localized process from which their actions emanate. This mechanism is shaped by historical factors such as education, culture, genetics, and other components that the agent lacks complete control over. These factors may influence the degree to which the mechanism functions, but so long as the mechanism remains functional – able to respond to reasons in general – then the agent, under my view, can be held morally responsible[2]. This should be the case regardless of how this mechanism was produced. Requiring the agent's values to track their autonomous construction would mean that other historical non-autonomous influencing factors, such as genetics and culture, which influence the construction of the agent's values, are moral responsibility undermining

[2] One argument in favour of a history-sensitive condition is offered by Fischer and Ravizza (1998), involving a drunk driver who, while intoxicated, strikes and kills a child (p. 195). Their argument is that although the driver lacks the mechanism at the time of the incident, the driver had the mechanism at the time they decided to drink. Given that Fischer and Ravizza find the driver intuitively morally responsible for the death of the child, they see history as a fundamental component of moral responsibility. However, I think it is quite possible to hold the driver morally responsible for drinking to the point of being so incapable, and not morally responsible directly for the death of the child.

as well, since they are not directly under the agent's control. If this were the case, then it would be an argument for incompatibilism, rather than a solution for compatibilism. This is because history itself would become moral responsibility undermining.

In order to make the preceding point clear, consider agents who were raised in ethically dubious environments and who manage to overcome that type of upbringing. If agents who are capable of morally-relevant control are exposed to the lies which informed their cultural upbringing, then they are capable, though not always likely, of being motivated to reject or leave their cultural or faith-based traditions. In contrast, some compatibilists, such as Susan Wolf (1988), find these types of cases intuitively responsibility undermining, since the influence of upbringing makes behaviour change unlikely. The problem with this is that even though change is unlikely, or even nearly impossible, the agent still maintains the capacity to respond to reasons. I see no reason for those who accept compatibilism to refrain from holding an agent, whose capability was shaped by upbringing, morally responsible. It should not matter, for those who find control to be moral responsibility conferring, whether the ability to endorse values is more or less difficult for differing individuals in differing situations, since that is trivially true about all agents who differ in ability and upbringing. Wolf's example of behaviour change being unlikely involves a dictator’s son named JoJo. Given JoJo’s upbringing it is likely that he will commit similar atrocities as those committed by his father. However, if it is possible that JoJo, upon being exposed to certain books or arguments, could change his attitudes and behaviours such that he fails to perpetuate the ill effects of his upbringing, then JoJo would seem to me adequately reasons-responsive and thus a candidate for moral responsibility. This remains the case even if JoJo is never exposed to those books or arguments. What matters is whether or not the *ability itself* – made possible by the reasons-responsive mechanism – is fundamentally undermined, and given that some individuals are capable of overcoming dramatic cultural indoctrination, it is not clear that cultures have such an ability undermining effect.

This is similar regarding neuro-manipulation cases since the manipulated agent, like the enculturated agent, is not likely going to change their behaviour. This means that if JoJo is indeed morally responsible, then there must be something about neuro-manipulation cases that differs such that neuro-manipulation cases are undermining. Mele (1995) thinks that the neuro-manipulated agent’s instilled values, compared to those of the enculturated agent’s, are inauthentic compulsions (p. 165-173). This means that the values were not autonomously constructed by the agent; not authentically ‘theirs’[3]. Dennett (2003) describes what Mele is proposing as the agent being “designed to endorse her own design” (p. 282). If the agent is designed to endorse their instilled values, then this seems to suggest that the manipulated agent’s reasons-responsive mechanism has been disabled. This is because it is not adequately responsive to incoming information. If the mechanism is disabled in the manipulation cases then there is no need to provide a history-sensitive condition. This is because the lack of moral responsibility is the result of the disabled mechanism itself and not because of a relevant feature of the agent’s causal history. If the mechanism is not disabled, then there is no relevant feature that differentiates JoJo from a neuro-manipulated agent or even from an agent who experiences what is considered standard moral education. All three of these agents maintain the ability to respond to reasons, although they differ in the sensitivity of their mechanisms given their differing histories. Every agent's mechanism differs in sensitivity, due to factors outside of their control, whether they are manipulated or not, so it is unclear why neuro-manipulation as opposed to cultural or educational manipulation is such that it should be treated differently.

In order for history-sensitivity to be practical for moral responsibility it must explain what feature differentiates manipulation cases from ordinary, everyday, value-engineering. Without an explanation of what this difference is, history-sensitive compatibilists risk having everything that plays a role in shaping an agent's development relegated outside of the agent's control. This means they remain vulnerable to Pereboom’s or similar manipulation

[3] I avoid here discussions of self-identity. It may seem that after an agent is manipulated, they lose something that was essentially 'theirs', something which is integral to their identity. However, it is not clear to me whether the self is such an essential quality. Rather than debate whether or not a manipulated agent is still the same agent on both sides of the manipulation, what interests me is whether the agent that remains is able in the reasons-responsive sense such that they – whoever 'they' really are – can be held morally responsible.

arguments, by conceding to the incompatibilist position that history can undermine moral responsibility. The only position available for compatibilists to stave off the creeping effect of the deep past is to bite the bullet, such that non-disabling manipulation, if possible, doesn’t matter for moral responsibility. This means to adopt an internalist position regarding moral responsibility. This is not to say that external factors do not influence the reasons-responsive mechanism, but that those external factors should not influence our ascriptions of moral responsibility.

Even if history-insensitive compatibilism is a reasonable position to adopt, it does not completely undermine the incompatibilist intuition that an action should not be produced by a deterministic process outside of the agents control (S). Although, I think it provides strong enough reasons for why S isn’t necessary. Whether or not one feels that S is intuitively compelling is not a satisfyingly robust reason to reject compatibilism, but there doesn't seem to be any other reason available. It is, therefore, still open that agents have the ability to control their morally-relevant decision-making processes under a history-insensitive compatibilism. But there is nothing to prefer this type of control, over the control provided by the daring soft libertarian account, beyond these differing intuitions. In other words, S is not necessary for morally-relevant control, since an agent's reasons-responsive mechanism is sufficient for free action, although some theorists find it intuitively compelling that their decisions be removed from a deterministic causal chain. Considering that history-insensitive compatibilism and daring soft libertarianism are both live options for control, the ultimate status of determinism is irrelevant to moral responsibility. This means that an agnostic account regarding determinism is a reasonable option for those who want to reject moral responsibility skepticism.

1.4 Agnostic Control and Concluding Thoughts

Fischer and Ravizza (1998), Mele (2006), as well as Peter Strawson (2013) hold views which are in some sense agnostic about the truth of determinism and indeterminism. They all agree that what matters most is that we retain moral responsibility, regardless of what the ultimate truth is regarding the physical nature of the universe. Fischer and Ravizza argue for what they call semicompatibilism, which is the idea that “moral responsibility is compatible with causal determinism, even if causal determinism is incompatible with the freedom to do otherwise” (p. 53). This type of ultimate control found in the freedom to do otherwise, or what Fischer and Ravizza call 'regulative control', is far too demanding given what is known about how the world works (p. 31-33). As Fischer (2012) explains

> ...my preliminary conclusion is that *if* causal determinism rules out moral responsibility, this is *not* in virtue of its eliminating regulative control (if it does indeed eliminate regulative control). This is an important point; I believe it is the 'moral of the Frankfurt stories,' no matter how they are told and re-told. Further, if this point is correct, it allows us to sidestep the traditional debates about the relationship between such doctrines as God's omniscience and causal determinism, on the one hand, and 'freedom to do otherwise' or regulative control, on the other. That is, we can sidestep these debates if we are simply interested in moral responsibility. Insofar as these traditional debates have issued in what I have called Dialectical Stalemates – black holes in dialectical spacetime – avoiding them may open the possibility of real philosophical progress. (p. 123)

Avoiding these “black holes in dialectical spacetime” is precisely what the agnostic wants to accomplish, and I agree with their goals. I especially agree with their goal to move the debate away from questions involving the ultimate state of the universe and towards more realistic conceptions of cognitive control.

Mele's agnostic autonomism is closely related to semicompatibilism in this regard. When pushed to explain whether he finds the manipulation arguments compelling, he admits it has an intuitive pull. For Mele, something about the determined histories makes it feel like the agent is not responsible. However, since Mele is more convinced of the need for moral responsibility, it is more likely that the agent is responsible than that determinism is responsibility undermining (p. 191-195).

Agnostic autonomists, like Mele, maintain that it is more compelling that agents have freedom then that agents do not. This means that if it is true that the universe is indeterministic, something more closely akin to daring soft libertarianism will help ground moral responsibility, and history-sensitive compatibilism will ground it if the universe is determined. Although I disagree with Mele about the acceptability of history-sensitive compatibilism, and prefer my own history-insensitive compatibilism, I think the position of agnosticism is the most compelling overall. Because of this difference with Mele, I call my account agnostic control, since it does not require that agents are entirely autonomously constructed. What matters is that agent's maintain morally-relevant control provided by their reasons-responsive mechanism. To be precise, agnostic control is a history-insensitive form of morally-relevant control that is agnostic about the truth of determinism, and is, I argue, the most realistic view of control available for those who seek to ground moral responsibility.

1.4.1 Conclusion

This chapter has been, to a large extent, an overview of the more recent history of debates regarding control, free will and moral responsibility and its relation to determinism and indeterminism. At the start I maintained that a coherent account of control must be possible, morally-relevant, as well as operated by the agent such that this decision-making process was 'theirs', in the sense that it is attributable to the agent. I attempted to distil the best accounts on each side of the debate in order to argue for what I believe is the best case scenario for morally-relevant control. Even though I ultimately think grounding moral responsibility will fail, I think it is important to stipulate the range of options we have to work with.

I have argued that noncausal and agent-causal accounts of control are not adequate, since they are incoherent. Event-causal incompatibilism is more robust in this regard, since it takes into account the need for an agent’s reasons to cause their actions, but leaves open the possibility of differing futures. Mele's daring soft libertarianism is particularly robust in this sense, since it gives agents control over the probabilities inherent in the indeterministic processes involved in decision-making. It is therefore a good candidate for an incompatibilist account of morally-relevant control. Pereboom's hard incompatibilism, the view that free will is incompatible with both determinism and indeterminism, represents the moral responsibility skepticism I wish to defend. His particular version of skepticism, however, does not succeed, since it relies on intuitions which are far too demanding, and are, additionally, not widely shared.

I further argued that history-insensitive compatibilism can be maintained. This in part is motivated by the fact that Frankfurt-style cases make the leeway (LW) criterion superfluous, rather than necessary for control. Secondly, the manipulation arguments against compatibilism fail to secure the need for the incompatibilist source (S) criterion, although it leaves it as a live option for incompatibilists. This means that there are two options available for the moral responsibility advocate to possibly locate the source of moral responsibility, since they are both coherent examples of our ability to have morally-relevant control. Furthermore, these options are similar enough that it makes the truth of determinism practically irrelevant for questions of moral responsibility. What should matter instead is not whether an agent is determined, but the extent to which the morally-relevant control available is sufficient to satisfy moral responsibility.

What often pushes someone to choose compatibilism over incompatibilism is their sense that incompatibilist intuitions feel more or less persuasive, but debates over competing intuitions are hardly decidable. Whether we choose compatibilism or incompatibilism there are still aspects of an agent’s deliberation which are not under their control. Agnostics are more compelled about the truth of moral responsibility than they are about the ultimate truth of determinism or indeterminism, and thus they have grounds for maintaining their belief in morally-relevant control regardless of what the ultimate truth of determinism turns out to be. This stance is what I call agnostic control, since its focus is on the nature of control not determinism. If agnostics are correct about focusing on control instead of determinism, then the debate about whether agents can be morally responsible depending on which state of physics happens to be true, is misplaced.

In the next chapter I will be providing reasons to think that agnostic control is undermined by luck. Considering that this chapter is an argument in favour of an account of moral responsibility it is important that I re-establish why this argument is necessary. My goals in this chapter are twofold. Firstly, I think the debate about moral responsibility is misplaced. Historically, and even presently (Pereboom, 2001; Harris, 2012), arguments in favour of moral responsibility skepticism have been motivated by determinism, which is an argumentative strategy I think is ultimately unsuccessful. The call, which I hope to meet in the next chapter, is for skeptics to provide a better answer to why moral responsibility is undermined. Secondly, I think a live option does indeed survive, regardless of the truth of determinism, which I call agnostic control. This makes the first goal more salient since the denial of moral responsibility should render all competing views unacceptable, but it also helps to focus the debate between skeptics and non-skeptics away from issues involving determinism and towards issues involving psychologically realistic accounts of control.

Chapter 2
Moral Responsibility Undermining Luck

The previous chapter was an attempt to argue in favour of a realistic conception of morally-relevant control, which is not concerned with the truth of determinism. This chapter will be addressing the contours of this ability; focusing on the morally-relevant aspects of agent's lives over which they have control, and the limits in their ability to hone this control such that they are capable of becoming morally better agents. Of particular interest will be the influence of luck, and the extent to which luck undermines an agent's moral responsibility by rendering various aspects of their morally-relevant control unfair. To the extent theorists want to argue for moral responsibility, they must account for the inequalities that exist between differing agents morally-relevant abilities, such that these differences in ability make moral judgments fair.

In section 1, I will define luck and fairness and their relevance in the moral responsibility debate. In section 2, I will explain the contours of our morally-relevant abilities, in order to address whether they are robust enough to survive luck objections. I will conclude that a particular aspect of luck is indeed worrying for a moral responsibility advocate. This aspect of luck, in particular a specific form of constitutive luck, involves the differing capabilities of agents, such that they render moral responsibility judgments unfair. In section 3, I address a response to my conclusion in section 2, such that the luck involved in each individual's differing capabilities allows for ascriptions of moral responsibility to be 'fair enough'. Those who advocate for the 'fair enough' position tend to compare morality to a game, but I argue that this analogy fails to recognize the differences between the goals of games and the goals of living a moral life or wanting a moral society.

2.1 Luck and Fairness

Neil Levy argues for a 'hard luck' account of moral responsibility (2011). The hard luck view is that moral responsibility is undermined, not by determinism, but by luck. I agree with some components of the hard luck account, although I disagree with Levy about some of the more nuanced moves in his argument. This is probably due to the fact that whereas Levy is a disappointed compatibilist (p. 2), I am a disappointed agnostic controlist. The use of the word ‘disappointment’ reflects the fact that moral responsibility is undermined by luck; however, I will give evidence in the next chapter for why this probably won’t be that disappointing. In order to address this difference between our views, I need to first define what is meant by luck in these moral responsibility cases and what it is about luck that can be responsibility undermining.

2.1.1 Luck

The problem with defining luck is that people often use it as a mere gesture, lacking any deep conceptual clarity (Pritchard, 2005, p. 125). In this sense, it has rarely been the subject of deep philosophical scrutiny. There are three main components of luck which seem to at least be implicitly agreed upon by those who have addressed this topic, chance, significance and control, although there remains disagreement in the details (Nagel, 2007; Pritchard, 2005; Levy, 2011). I will take each component in turn in order to elaborate the extent to which luck could be moral responsibility undermining.

Chance is the first component that usually comes to mind when people think about luck. The best example of this is a lottery, since, if the odds against an individual winning are incredible, then the person who wins is lucky. In other words, the chance of a single individual losing is so high that if they ultimately win, that event is lucky for them. Although chance is important, whether something is ultimately lucky for the agent is moderated by the agent’s control and the significance of the result for the agent.

Chance is partly what motivates the intuitions for and against compatibilism that were discussed in the previous chapter. If everything is determined, then there is no chance an agent could have done otherwise, or the opposite concern that if everything was up to chance there is nothing an agent could do about it. In the case of determinism, even though the odds of winning would remain the same, the winner would have been destined and thus there would exist no chance in this universe. If this is the case, then it would seem that the debate over determinism does indeed matter. However, Levy (2011) offers a modal account which I think is particularly apt at describing chance in a deterministic universe. Levy describes this account as follows:

> Event *E* is chancy if it occurs in the actual world at *t**, but it fails to occur in a large enough proportion of possible worlds obtainable by making no more than a small change to the actual world at *t*, where *t* is a temporal interval just prior to *t**, and the agent lacks direct control over *E*'s occurrence. (p. 19)

This modal account is not meant to imply that there exists actual chance in a deterministic world, but it is a possible way of conceptualizing how different worlds could have turned out. Seth MacFarlane, the creator of the cartoon television show *Family Guy*, was chancy lucky in this sense. He was late for a flight which ended up being hijacked the morning of September 11th 2001 (Weinraub, 2004). Being both hung-over, and having his agent misinform him that the departure time was at 8:15 and not 7:45, MacFarlane arrived too late to board his flight. If the universe is determined, then there is no chance that MacFarlane would have done anything differently in the actual world, but considering how easy it is to conceive of possible worlds in which MacFarlane arrived for his flight on time, for example, a world where his manager told him the correct departure time, MacFarlane was indeed chancy lucky given this modal account.

There is also a sense in which luck can be non-chancy, such that the distance between possible worlds is too far to be chancy. Constitutive luck, which is described by Thomas Nagel (2007) as the luck which factors into who the agent becomes, can be a luck of this kind (p. 297). This is to differentiate it from circumstantial luck, which is more akin to the luck experienced by Seth MacFarlane in the previous example. Agents lack control over the historical and cultural processes which were constitutive of their morally-relevant decision-making processes, and it is not clear how the universe could have been easily different such that an agent's constitution would not be the same. What makes this a form of luck is the extent to which agents vary from most other human beings. As Levy (2011) explains "An agent may be lucky to be clever, but we would not say that an agent is lucky to be cleverer than my dog” (p. 33). This is because most agents are cleverer than dogs. The more these traits vary among the reference group – human beings – the more the agent is lucky or unlucky for having the particular traits they do.

The next component of luck is significance. It is important to note the asymmetry that exists in the ascription of luck. Although an individual who lacks some kind of ability might be unlucky, it is not clear that someone with that ability would be lucky. For example, someone who is more forgetful might be unlucky, while someone with an average memory might not be lucky or unlucky. This applies to chancy luck as well, since losing the lottery doesn't make someone unlucky given that they were probably going to lose in the first place. This means that the significance of the event moderates the extent to which luck is ascribed, when compared to the proportion of nearby possible worlds in the case of chancy luck or compared to the proportion of individuals in the reference group which end up having a particular disposition in the case of non-chancy luck.

It is also important to note that, in the case of luck, significance can moderate the extent to which the chancy or non-chancy factors matter for the luck claim. In a game of Russian Roulette, where a participant places a bullet into the cylinder of a gun, spins it, and then pulls the trigger on themselves, the significance of the event renders the odds of the outcome less important. If the odds of surviving are in the agent’s favour, such that the odds of dying are 1 in 6, it is still lucky if they survive, and unlucky if they die, given how significant life and death is. Compare this to how insignificant winning a one dollar bet would be with the odds of success being 5 in 6. This is important for moral cases since it could render the small differences between agents incapable of evening out, given the high stakes involved in being praised or blamed. Having a lucky slight advantage over a fellow agent could make it unfair to treat each agent equally.

The final component of luck is control. In some circumstances events become less lucky through the application of skill. In other words, an event is lucky the more an agent lacks direct control over it. As Nagel (2007) explains “it is intuitively plausible that people cannot be morally assessed for what is due to factors beyond their control” (p. 295). In the case of the lottery, an agent is lucky because they lacked control over the results. In fact, lotteries are designed to remove direct influence by blinding and randomizing the selection process, at least to the extent that people can't easily ‘game’ the system. In this sense they are designed to remove control from the agent. If instead of winning the lottery an agent earned their million by working hard, then their earning that million is not necessarily lucky, since the agent, in some sense, though not completely, had control over the outcome.

Technically, for the lottery, an agent could increase their chances of winning by buying extra tickets. Thus, in some sense, like the chance involved in Mele's daring soft libertarianism, agents could have control over the relevant probabilities. Mele (2006) gives the example of Ann, who promises a friend that at noon she will flip a coin. For the most part Ann is compelled to flip the coin, because she feels that people who break promises are bad, she likes her friend, etc. However, the only thing that would count as a reason for Ann against flipping is an extremely negligible desire to break a promise. Now, it is not likely that Ann will not flip the coin at noon, but given the chance involved there is a possible world in which Ann does not flip the coin. In some of these worlds Ann would fail to flip the coin at noon not because of any of her intentions or reasons, but because of accidental features of the environment, which at present prevent her from flipping. For example, an atom in the wall, due to quantum indeterminacy decayed in this world, as opposed to the world where Ann successfully flipped the coin, and caused the roof to collapse preventing Ann from flipping the coin. The only morally relevant worlds are ones in which Ann intends to not flip the coin for reasons, and that the indeterminism present in the situation is internal to the agents intending. Mele refers to these worlds as *N*-worlds. If Ann performed the noon time coin flip (*C*) in this world (*W*) then in some other *N*-world Ann does not *C*. This is a problem succinctly captured by Mele:

> Pick an *N*-world *Na* in which Ann decides at noon to hold the coin for another minute and continues holding it. Her making that decision at noon rather than *C*-ing, as she does in *W* - that

> difference – seemingly is not accounted for by anything. *Na's* divergence from *W* begins with Ann's deciding at noon to hold the coin. So there is no difference between *W* and *Na* to account for the difference at issue... If nothing accounts for the difference, the difference is just a matter of luck. (p. 59)

If, however, Ann's decision, which was the result of her skill as a moral agent, influenced the probability that the more desired outcome occurred, then according to Mele, the outcome was not *just* a matter of luck (p. 114). The fact that it isn't just a matter of luck means, according to Mele, that it is still possible to hold agents morally responsible.

Levy (2011) thinks this is insufficient at ameliorating luck, since the agent lacks direct control over the event itself (p. 19). According to Levy, luck, in the sense of lacking control, is all pervasive. Given that control is necessary for moral responsibility, this would mean that luck, by virtue of undermining control, makes moral responsibility impossible. Since luck is all pervasive, in that it applies to every aspect of life, there is no way that moral responsibility could be possible at all. This concern is also raised by Nagel (2007) “If the condition of control is consistently applied, it threatens to erode most of the moral assessments we find it natural to make” (p. 296). Although I agree with Levy that Mele's argument fails, I disagree about why this is the case. I think, contrary to Levy, that having some control could matter and that demanding ultimate control over the situation is demanding too much. It is trivially true that agents lack control over all the historical or cultural processes which happen to impact who they become. This includes the contents of their deliberations, as well as lacking control over the more chancy factors which do the same. But so long as agents remain capable of having *some* control over their morally-relevant decision-making processes it is possible that agents have *enough* control to be morally responsible.

Cultural influences may render certain behaviours more likely, such as the honour traditions enculturated in the southern United States (Cohen, Nisbett, Bowdle, & Schwarz, 1996). But these influences are provided by the societal structure itself and do not emanate from the agent’s capabilities provided by their reasons-responsive mechanism. As I argued in the first chapter, the reasons-responsive mechanism is the most appropriate place to locate blame given my preference for history-insensitivity. Therefore, these cultural influences do not undermine moral responsibility completely, although they do hinder the agent's choices currently. What matters, with regard to luck and moral responsibility, is to what extent there are differences in abilities among agents in the morally-relevant reference group. If an agent’s differences in ability are, in some sense, under their control, then I think it is possible to hold those agents in the reference group morally responsible. However, if, due to luck, agents are unable to alter their abilities, in such a way that it levels the playing field, then this kind of luck can be moral responsibility undermining. My argument for this will be made more precise in section 2.

2.1.2 Fairness

The main reason why luck is worrying for moral responsibility is our sense of fairness. Fairness is the idea that people should treat others equally for things that are relevantly similar, and differently only when there is a desert-based reason for the different treatment. It would be unfair to judge someone based simply on the colour of their skin. To do so is unfair partly because it does not play a role in their abilities, but also because skin colour is something an agent lacks control over. There is also the contrastive fact that the differences between those with differing skin colours are not morally-relevant, and should not count as reasons to treat these individuals differently from one another. This desire to treat like cases alike is grounded in a sense of distributive justice. As Samuel Freeman (2007) explains “A primary role for a principle of distributive justice is to provide standards for designing, assessing, and publicly justifying the many legal and economic institutions that structure daily life” (p. 450). This view, which shares adherents such as David Hume and John Rawls, perceives justice, as fairness, as an integral psychological component which accounts for the non-oppressive stability of societies (p. 260). Fairness is also considered by moral psychologists to be an innate component of moral reasoning that has been a key component of our social evolution as a species (Haidt 2012, p. 130). Fairness, then, is an integral part of what it means to be human.

Luck is by definition not under the agent’s control, and it impacts all aspects of their lives including the colour of their skin and the factors which influence their moral deliberations. However, as argued previously, not every example of a lack of control is moral responsibility undermining. Even if significant aspects of life are not directly controlled by agents, some of life is, and some control, rather than complete control, could be enough to make moral responsibility possible or at least ‘fair enough’. What makes moral responsibility unfair is the extent to which we treat all individuals equally when it comes to desert-based distributions, even when an individual’s ability to control is significantly different from other agents. This means that the kind of luck that is moral responsibility undermining is contrastive as opposed to luck being universally undermining as argued for by Nagel and Levy. For those who argue in support of moral responsibility this may seem like a small problem, since there are obvious cases where we treat individuals who have committed the same type of crime differently based on their mental acumen and that the differences that remain between most agents is too small to matter. My concern, which will be addressed in the following sections, is that these differences are more prevalent than we’d like to think, and even small differences in the range of agents abilities can make moral responsibility unfair or at the very least not fairly tenable.

If, as Levy (2011) argues “preserving moral responsibility comes at the cost of fairness” (p. 10), then one way to defend moral responsibility is to, in some sense, argue that fairness is not that concerning. For those who defend moral responsibility, arguing that fairness is not concerning would be absurd. One of the main reasons for keeping moral responsibility is the fairness in our desert-based distributions; otherwise there would be no purpose in holding people morally blameworthy or praiseworthy. This is because differential treatment, in the sense of distributive justice, must be justified based on reasons which distinguish why this differential treatment is necessary. Desert would count as one of these reasons, but we wouldn't need desert if the idea of distributive justice is not maintained. This means that discarding fairness in order to defend fairness is obviously problematic given the contradiction it presents. It should, therefore, be a motivating concern for those who wish to defend moral responsibility to determine whether or not our moral judgments are indeed fair, or at least ‘fair enough’. If they cannot meet this standard then moral responsibility skepticism is an appealing option.

2.2 The Contours of Control

Now that it is clear what luck and fairness are, and the work they are doing in this debate, I will begin to address what aspects of control must be undermined by luck in order to undermine moral responsibility.

There are several aspects of an agent’s life that they absolutely cannot have control over for obvious reasons. For example, agents cannot control historical events which occur prior to their existence. Agents also lack any control over their genetic heritage which is provided by their parents. Agents owe a great deal of their existence, including their moral faculties, to these cultural and biological influences, and yet have no way of controlling them. Demanding that agents have control over them, in order to be morally responsible, is simply demanding the impossible. It is not clear that this demand is intuitively appropriate. There is a way in which these historical facts can indeed render some individuals not responsible, but this depends on the extent to which the historical facts leave the agent unable to maintain morally-relevant control. However, if certain historical facts leave some agents more or less capable than other agents, then I argue that this inherent inequality can render moral responsibility practically unfair.

In order to make this case, it is important to clarify precisely how this control ability manifests. In the previous chapter I discussed the morally-relevant decision-making processes which represent our ability to control, but I did not explain how this control functions. In particular I did not explain how agents process the content of their deliberations, such that they can be held responsible for their processing. Part of this explanation is going to revolve around what the 'agent' is that is capable of this kind of control. Without trying to assert anything too controversial, for the sake of my argument it is enough that I adopt the working notion of the ‘self’ used by most neuroscientists and psychologists; that the self is the subject (loosely defined) of experience and is intimately connected with the brain

(Blackmore, 2012; Thagard, 2010). More specifically agents tend to align themselves with their conscious experience, such that whatever occurs outside of their awareness is outside of their control and thus not really 'theirs', at least not in the moral sense (Levy, 2014). Levy calls this the 'consciousness thesis', which states that we are morally responsible “only when our consciousness of the facts that give our actions their moral significance are those actions expressive of our identities as practical agents and do we possess the kind of control that is plausibly required for moral responsibility” (p. 1). For example, alien hand syndrome, as the name would imply, is when the activity of the hand does not feel like it belongs to the owner, since the behaviour happens outside of their control and awareness (Wegner, 2002, p. 4-6). People who suffer from alien hand syndrome will find themselves reaching for things, or opening things, even when they don't want to perform those actions. Though the hand belongs to the agent, and the syndrome has a neurological source in their brain, the behaviour is not in the relevant sense 'theirs'. When an agent’s conscious experience aligns with their experience of control, such as the sense that an agent’s deliberations cause or help to bring about future actions, only then do they tend to feel like they are in control of their actions.

Missing from this picture is the sense that people view themselves as a unified whole, an ego which has executive power over the contents or their deliberations. Even though this is how some people feel, agents are not the conductor directing the neuronal spikes in their own brain. Nor are they the file clerks diligently searching folders for the right memory or bit of information to help solve a problem. 'Self' is merely the abstract label applied to the neuronal processes which factor into an agent’s conscious experience. This has led many theorists to call the self an illusion, but this does not mean the self, in abstract, does not exist (Hood, 2012). Like a card trick, the self is merely not, in fact, what it seems to be. Many people find this view of the self to be unnerving, in that it undermines our *special* ability to control and thus be morally responsible. But as I quoted Dennett saying in the first chapter, of course *that* kind of control (ultimate control) doesn't exist. The self at the center of experience, the unified 'I', doesn't have control over anything, but an agent’s brain, from which this unification emanates (or that is co-extensive with this unification), still possess the abilities and properties which allow it (I/we) to process morally-relevant concerns. But even this type of unified abstract self, or the deliberative processes which this self represents, does not have executive control over the contents of what is being deliberated. So even though, in some sense, this abstract self has control, in that the processes of the brain which are experienced consciously are in a relevant sense the 'agent’s', there are still some features which factor into this control which the self lacks control over. What should concern us here is not whether or not individuals are special ‘selves’ capable of ultimate control, but whether the lack of control over the features which factor into the shaping of and the content of an agent’s deliberative processes matters for moral responsibility.

Levy (2011) argues that this type of control over the contents of an agent’s deliberations is necessary for moral responsibility. As he explains, the type of control which allows for free will requires:

> (a) psychological control over our actions, such that if our mental states rationalize a prospective action (where to rationalize an action is to make that action the best alternative open to the agent, from his or her own point of view), if the agent act intentionally at all, the agent will perform the rationalized action; and (b) control over the psychological states that rationalize actions, such that if an action is rationalized by the agent's psychological states, the agent is blameworthy for that action only if he or she is blameworthy for acquiring or retaining the psychological states that rationalize it. (p. 5)

Given that agents lack control over (b), the historical features which factor into the psychological states which help to rationalize their actions, agents should not, according to Levy, be treated as morally responsible. This is because historical luck is all pervasive given that the control necessary for moral responsibility is epistemically demanding (p. 110-132). This means that because agents often behave how they think they should, in order to be responsible we have to blame them for what they should have known, but requiring that agents be responsible for their ignorance seems intuitively problematic. As Levy explains “Since in these common cases agents are acting as they believe they ought, we can apparently find them blameworthy for their actions only if we can blame them for their false beliefs" (p. 131).

I think this is the wrong place to locate where luck is responsibility undermining. This is partly because I am sympathetic to the history-insensitive control based moral responsibility views that Levy finds unconvincing. Dennett (2014) argues against these types of epistemic concerns by reflecting on whether or not the right play will occur to a bridge player,

> *Contrast* your competence at this moment with the 'competence' of a robotic bridge-playing doll that *always* plays its highest card in the suit, no matter what the circumstances. It wasn't free to choose the six, because it would play the ace *whatever the circumstances were* whereas if it occurred to you to play the six, you could do it, depending on the circumstances. (2014)

In other words, given the circumstances, some control remains in principle. There is a sense in which such remaining capabilities could be used in support of justifying moral responsibility, if the ability to control is equally available to all individuals included in the morally-relevant reference group. This does not mean that moral responsibility requires that all agential actions be identical, since the differing historical factors, which influence the sensitivity of the agent's decision-making processes, still allow for differing responses between agents. However, if the agents in the reference group are all equally capable, then they would still be capable in the same morally-relevant sense. This would mean, given that these historical features are out of *everyone's* direct control that luck would be diffuse. If all agents equally lack control over something, then it is not a concern for fairness.

Levy does not think that the kind of control which remains, in principle, matters, as he explains "I don't have a kind of ersatz control over my car if the steering wheel falls off; the fact that cars are *in principle* controllable does not alter my lack of control in that particular circumstance" (p. 207). There are two responses to this kind of claim. The first is that it seems to suggest that moral responsibility requires sameness, such that if agents only had and lived identical lives as everyone else they would be the types of agents who are candidates for moral responsibility. As someone who shares Levy's moral responsibility skepticism I don't find this overly problematic as an intuition, except that it probably wouldn't convince anyone else, especially those who see these kinds of differences as being 'fair enough'. The second response is that the analogy between the steering wheel falling off the car and the historical influence of available psychological states is not quite tight enough to be convincing. This is partly because the steering wheel is the precise connection between the driver and the car which serves as the mechanism in virtue of which the driver has control. In the case of deliberation agents are simply ignorant of a few reasons which could have altered the decisions they eventually arrived at, although the capacity to control was maintained throughout. A better analogy would be to suggest it is like a moderately essential component of the car malfunctioned leaving the driver capable of driving the car with a chance of an error (false belief), but this doesn't have the same edge to it as severing an agent’s ability to control *tout court*. A more convincing argument is to find fault in the contrastive fact of each individual’s differing abilities, since, I argue, it makes more salient the fairness concerns that Levy and I share.

In order to highlight the contrastive fact which undermines moral responsibility it is important to define the contours of the ability itself. Some abilities are of the type that there is nothing an agent can do about them. Aside from the theoretical possibility of neurological or physical alterations, some of an agent’s abilities cannot change. For those who lack the ability to do *x,* they represent the floor of that ability. It is not possible for them to be any less able than they are at being able to do *x*. For those who are able to do *x*, there are several possible distributions of ability. One possibility is that doing *x* is an all or none phenomenon, such that if a member of the relevant reference group, in this case all human beings, is capable of doing *x* they are no more or less capable of doing *x* than other capable members of the reference group. Any move away from the all or none distribution means that there will be variation among the reference group, consisting of individuals who are more or less capable of doing *x*. However, given that I stipulated this ability is of the type that cannot change, there is nothing an individual could do to alter their place in the distribution, there is no way for individuals to become capable, or to become more or less capable than others. It is also relevant to mention that there is the possibility of a ceiling effect, such that beyond a certain level of ability to *x*, it is not possible to be any more capable of being able to do *x*. Everything between the floor and the ceiling consists of the range of options available to do *x*.

If the distribution is such that an agent being able of controlling their morally-relevant decision-making processes is an all or none phenomenon, then moral responsibility could be fair, especially if we only hold those who are *capable* as candidates for moral responsibility. However, the further this distribution gets from all or none, the more weight seems to be given to the fairness intuition. If agents are not capable of making their own morally-relevant decision-making processes more apt to deal with moral situations, then it seems harsh to judge all individuals of the reference group equally, even if those differences could be small. This is because of how moral responsibility, and the rewards and punishments that are justified by it, are significant to the agent, thus requiring the satisfaction of a more robust form of fairness.

It is possible to include meta-abilities, such that individuals could be able to make themselves better able to do *x*[4]. An easy way to represent a meta-ability is to contrast, for example, being able to drum, from the ability to make yourself, through training or practice, to become a *better* drummer. The introduction of the meta-ability allows for a more overlapping distribution of ability, which could possibly even out the morally-relevant playing field. Imagine that each individual's ability is represented by a bell curve[5], such that the x-axis represents the moral valence of the action and the y-axis represents the agent's

[4] There is a worry here of a possible infinite regress, such that an individual is meta-meta-able if they are able to make themselves better able to make themselves better able to do *x*. However, given that this ability is reflective of specific and finite cognitive processes there is a possible end to this kind of series given the evolutionary modularity of cognition and I think the appropriate cut-off for my purposes here is the meta-ability.

[5] It is important to stress that this use of the bell curve in this example is to help illustrate the kind of phenomenon I have in mind and is not necessarily indicative of how human capabilities are actually distributed.

ability to control their morally-relevant decision-making processes. The more you move from the right or to the left of the center of the bell curve the less able the agent becomes at making morally worse or better decisions. That means there exists a tight range of more likely moral decisions for each agent, with decisions that are less likely further to the right and left. The meta-ability would be the ability to move the center of that bell curve, which in this example represents an agent's more stable dispositions or character traits, and possibly move the extremes, such that the agent can become more likely to make morally better decisions. If one were to graph this meta-ability, such that the x-axis would be the representation of how able an agent is at controlling their morally-relevant decision-making processes and the y-axis would be the representation of the meta-ability, then there would exist a range in which agents could improve. If you were to plot all individuals of the reference group on the same graph there will be an extensive amount of overlap in the meta-ability. The only way for this distribution to be fair, in terms of holding people morally responsible, is that even though each individual's peak would be different the extremes at the left and right would be the same for everyone. In other words, although individuals will differ from each other on average, they remain equally *capable*, though more or less likely, of making the morally worst and best decisions. This phenomenon is best captured by JoJo, who, given his upbringing, was more likely to commit the same atrocities as his father. However, as I explained in the previous chapter, if the reasons-responsive mechanism is still functional it is possible for JoJo to change if he is exposed to the right arguments. Thus he remains capable of making morally better decisions, although, given his life history he is less likely to be exposed to the precise factors which motivate him towards making better decisions.

One reason to think that this meta-ability is insufficient for moral responsibility is the extent to which this distribution accurately represents the existence of such a shared meta-ability. Part of the problem is that there is no easy way to assess whether an individual is fully capable of making these extreme decisions. Dennett (2003) thinks that this might mean "that there were fewer morally responsible people in our society than we had heretofore supposed" (p. 292). However, the smaller the reference group gets the more impractical the implementation of moral responsibility becomes. It would be arguably problematic for society to be set up such that there were different moral leagues, where rewards and punishments were doled out differently, much like we reward the relatively small number of expert baseball players with higher salaries. Again, this is because of how unfair such a system would be, given the constitutive luck which created such differences in the first place. It is ironic that Dennett asks us to contrast our ability to that of a bridge-playing doll in order to show how humans have a superior ability, and yet it is precisely this contrast which I think undermines his view. This is because the difference between our abilities and that of the dolls is only a matter of degree, and this remains true between humans, although the contrast is less stark. Furthermore, even if there would be some obvious cases of agents who lack this capability and who could be removed as candidates from the moral responsibility reference group, the vast majority of individuals in the moral responsibility reference group would exist with psychological states currently too chaotic for us to accurately assess the degree to which they could be considered equal in their ability to alter their range of moral decisions. However, even though it is too chaotic for us to have the precision to determine that agents are equally capable I think there are still reasons to think agents lack this shared ability.

There is an obvious sense in which agents do vary in ability, in the sense just discussed, and that this variation is the product of non-chancy constitutive luck, specifically in an agent’s development. This variation is an inherent quality of the messiness of our evolutionary development (past and present), and there is no reason to think that an agent’s ability to control is such a robust evolutionarily stable strategy that it lacks variability among the agents lucky enough to have it. To refer back to the drumming example, I have never been able to drum well, not for lack of trying. However, I know several drummers who, in some sense, were initially better at it than I was. Possibly they were born more coordinated, or had a better sense of rhythm. I may be able to improve at drumming, but no matter how hard I try I will never reach the ability, of say, Neil Peart. When cyclist Michael Shermer (2001) was asked about how he felt placing third in a race, he responded “I should have picked better parents” (p. 92). This isn't to say that training and environment do not matter, but that given our starting positions, some of us are more or less capable than others at being great drummers, great cyclists and, by extension, better moral agents. Even if the differences of ability between moral agents are small, given the significance of a system of praise and blame, small differences matter. Just consider that those small differences could allow for one individual's upward mobility and another individual's imprisonment.

2.3 ‘Fair Enough’

One argument against my claim that moral responsibility is unfair is that asking for the moral responsibility system to be fair is asking for too much. After all, life isn't fair. What is important is that we try to *make* life fair as best we can; that life, in the end, can be 'fair enough'. Dennett (2012a) raises this objection by comparing the system of moral responsibility with the rules of games, showing that the rules are designed to make the game as fair as possible. His argument is that insofar as we need a system of moral responsibility, civil society and its institutions exist to diffuse the effects of unfair starting positions “by minimizing the amplification of advantage and disadvantage that otherwise would probably occur” (2012a).

What makes the game analogy particularly apt for Dennett's comparison is that games come with a set of punishments and rewards. A player who makes an illegal move can be penalized by deducting points or by preventing them from being able to play for a few minutes, or they could even be thrown from a game completely. Players are also rewarded by performance through higher salaries and benefits. Even though the agents involved lacked a fair starting position, and the rules were in some sense arbitrarily designated, most players accept the rules of the game as fair. As Dennett explains, this is because “our game *draws* some sharp boundaries in order to *create* an environment in which desirable activities can take place” (2012a).

Dennett's main point, as it relates to moral responsibility, is that since nothing is perfectly fair, the moral responsibility skeptic needs to address whether or not a slightly less than perfectly fair system could ever be devised, such that the agents involved would agree to participate. Dennett thinks that such a less than perfectly system does exist, the moral responsibility system, and barring the possibility of something better, Dennett thinks the moral responsibility system is the best game in town. He also thinks that most people would agree to play the game. In other words, people will agree to be candidates of praise and blame. Furthermore, Dennett's (1984) view is that “Instead of investigating, endlessly, in an attempt to *discover* whether or not a particular trait is of someone's making – instead of trying to assay exactly to what degree a particular self is self-made – we simply *hold* people responsible for their conduct (within limits we take care not to examine too closely)” (p. 164).

Bruce Waller (2011) argues that this take-charge aspect of Dennett's view means closing our eyes to the details that could possibly exonerate individuals:

> The sad consequence of that moral responsibility myopia is that the conditions that foster and undercut the effective exercise of take-charge responsibility must be ignored in order to sustain the moral responsibility illusion. If we want people to take responsibility – take-charge responsibility, the only kind they can take and the only kind worth taking – then it is essential to look carefully to the cultural and the psychological conditions that foster (and often impede) the effective exercise of take-charge responsibility. (p. 111)

I agree that this is essential, since without looking too closely we give up the goal of fairness, even the goal of being 'fair enough'. Dennett responds to this criticism by again utilizing the game analogy. He thinks theorists like Waller and myself are suggesting that “We should eliminate the rule in baseball that distinguishes unfairly between a 350 foot home run and a 348 foot fly ball. We should look at each struck ball very closely and stop this arbitrary distinction between hits and fly balls, dealing with each feat in all its natural, real properties” (2012a). However, this seems to assume that the arbitrariness, when it comes to the rules of games, is in some sense equivalent to the arbitrariness as applied to society’s laws. The legal case plays a more significant role in our lives than whether or not the boundary of a home run is moved by a few feet. And I would hope this significance is reflected in a more precise and detailed rule constructing process which pays close attention to all the morally-relevant factors.

There are two aspects of the game analogy I will address that make settling for 'fair enough' worrisome. The first has to do with fairness within the game itself and the second with the ability of agents to choose to play the game in the first place.

2.3.1 Fairness Internal to Games

The game analogy fails to take into account that we don't put novice players up against the experts. As Tom Clark explains “Games are primarily about ensuring fair competition between more or less equally talented players to see who's best; we don't put novices up against experts and expect a good game” (2012). This is in part because the goal of a game is to determine who the best player is. Since novices are generally not the best players, the novices will most likely be beaten by the experts. Given the goal of games that kind of match up would be superfluous, let alone boring to watch. Often the rules of play are designed either to make specific attributes more salient, such as determining the fastest runner, or by increasing certain chancy elements in order to decrease predictability, which happens to increase fun, such as the fun in playing slots. Structuring society such that those individuals by virtue of constitutive luck are made the winners of society, clashes with the intuitive sense of fairness represented in the ideal of distributive justice. I also doubt that many would appreciate the government making life more unpredictable to increase our happiness. Not knowing when your next meal will come is not as fun as not knowing what card you will draw next in a game of poker.

Dennett agrees that changing the system matters, such that we should actively model society in ways that the playing field becomes more level, especially considering the problems with our current criminal justice practices that he, along with moral responsibility skeptics, wishes to reform (2012a). However, it is not clear how changes in rules affect the extent to which, in the case of baseball, someone could become a better hitter. Similarly, adjusting the rules and regulations which factor into the distribution of rewards and punishments is not going to change the fact that some agents are more capable moral agents than others. Given that some of the difference in ability is not amenable to change it is not clear to what extent rule changes can level these disparities in ability such that the outcome would be fair.

The second reason the analogy fails, which has been indirectly mentioned, is that the stakes for life are far higher than the stakes for games. A loss at baseball is not without a cost, but the cost of being socially disadvantaged by an unlucky constitution is far worse. Games may be 'fair enough' in that most people agree to play them for fun and for challenge, which demands a level of fairness, but not complete fairness. After all, there is a certain excitement to beating another player who is technically better, hence, why some people tend to cheer for the underdog. This can make playing slightly unfair games more worthwhile, than if the competition was identical. But it is not clear that people would agree to participate in an unfair society, which distributes rewards and punishments based on luck, unless they knew ahead of time that they were going to be in the place of advantage. This is what motivates the Rawlsian conception of fairness idealized in his hypothetical original position (Freeman, 2007, p. 143). From which, blind to how they will turn out, agents determine the structure of a fair society. Although some aspects of Rawls' approach have met with considerable criticism, it is enough for my purposes here to propose that not everyone would agree that the system of moral responsibility is 'fair enough'.

2.3.2 Fairness in Choosing to Participate

If Dennett is right, that the moral 'game' is internally fair, to what extent is our choice in participating in the 'game' fair in the realm of moral responsibility when compared to a sport like baseball? Dennett explains that part of the reason people participate in the game of baseball is because they consent to the rules of the game prior to playing. Basically, that if you don't like the game, you don't have to play it (2012a). Dennett asserts “You don't *have* to play the moral responsibility game; you can be a hermit on an otherwise deserted island, fishing and foraging. But if you want to enjoy the benefits of living in a civilized society, you have to play the game” (2012a). I agree, that compared to the state of nature, civilization along with the moral responsibility system have been a net benefit for human beings in general. However, I do think there is room for improvement, and I think the worldview implicit in the moral responsibility system is an impediment towards further change. That being said, it is not clear to what extent there is an agreement made by the agents within the moral responsibility system as to their consent of the treatment so delivered. It is also not clear to what extent agents have the ability to escape – become hermits as Dennett phrases it – such that they can live under a different system of governance, or according to the baseball analogy, to play a different game.

What this argument seems to hinge on, is the extent to which agents, as participants in the system, tacitly agree to be treated as morally responsible. When left with the option of being a hermit or a member of society, agreement seems more like a form of coercion than the result of an agreement. Referring to becoming hermits, Waller responds

> For profoundly social animals, that's not an attractive option – especially when those who remain in the game would classify us as demented, and fit only for treatment, and unworthy of respect or affection. If we don't play we are banished not just from society, but from the human community. (Given that alternative, it is hardly surprising when even those who are most severely mistreated -- by a U.S. system Dennett agrees is grossly unfair -- insist that they want to remain in the game.) (2012)

Dennett sees his participation in the moral responsibility system as not a coerced decision, but a reasonable one given the alternatives. He even suggests that remaining in the currently available system is a "wonderful bargain" while still admitting there are issues involving criminal justice reform that need attending to (2012a). However, most individuals did not have the luxury of being constituted as the brilliant, morally adept individual that Dennett happens to be. They therefore lack some of the options that are available to Dennett, and might only have to choose between either a life of painful solitude or a life on the receiving end of society's moral indignation, which is not that wonderful of a bargain all things considered.

Dennett's argument, which sounds vaguely like an appeal to tacit consent, has a history of being criticized. A great example of why tacit consent is absurd was provided by David Hume (1777),

> Can we seriously say, that a poor peasant or artizan has a free choice to leave his country, when he knows no foreign language or manners, and lives from day to day, by the small wages which he acquires? We may as well assert, that a man, by remaining in a vessel, freely consents to the dominion of the master; though he was carried on board while asleep, and must leap into the ocean, and perish, the moment he leaves her. (Mil 475)

Similarly with moral responsibility, it is not clear in what sense an agent consents to the practice of holding people morally responsible if their only alternative is the equivalent to jumping into the ocean and perishing.

2.4 Conclusion

In this chapter, I argued that constitutive luck impacts the extent to which agents differ in their ability to control their morally-relevant decision-making processes when compared to other human beings. This difference in ability, which renders some individuals incapable of reaching certain moral heights or depths, renders moral responsibility judgments unfair, given a ceiling or floor effect which prevents agents from further improvement. I argued that even if these differences are small, given the significance of being held morally responsible, such that agents are now capable of being subjected to various forms of praise and blame, moral responsibility remains inherently unfair.

I then defended my moral responsibility skeptic account against the criticism that the moral responsibility system is 'fair enough'. Given the disanalogy between games and society, I think there is reason to expect a higher degree of fairness than that provided by the moral responsibility system. In other words, I don't think it is necessary for us to settle for 'fair enough'. Whether there is, in fact, a better option than moral responsibility is going to depend upon whether moral responsibility is in some sense practically necessary, such that without it life would be miserable. I will now address that claim in the following chapter.

Chapter 3

Reactive Attitudes and Criminal Punishment in Relation to Moral Responsibility Skepticism

In the previous chapter I argued that moral responsibility is unjustified given the constitutive luck which helped to shape each individual's varying morally-relevant abilities, making the practice of holding people morally responsible unfair. My task in this chapter will be to address several arguments which claim that the practice of moral responsibility is in some sense practically necessary even if it is not fair, and thus moral responsibility should not be rejected. There are two areas which have the most support among those theorists who find moral responsibility skepticism uncompelling; first that moral responsibility is a natural extension of an agent's participant reactive attitudes of praise and blame making the abolishment of moral responsibility not likely or in some sense not desirable. The second is that the current system of criminal punishment which is defended by our practice of moral responsibility is necessary for the well-functioning of human societies. I will argue that both these positions fail for similar reasons. In order for moral responsibility to be practically necessary the supposed benefits must be shown to depend on the practice of moral responsibility. They also must be shown to indeed be benefits. If I can show that the supposed desirable effects of moral responsibility are not dependent upon moral responsibility or that the supposed benefits are in fact not benefits, then the practical necessity of moral responsibility is not warranted. This would mean that, all things considered, moral responsibility skepticism would be a reasonable position to adopt.

In section 1, I will address our natural propensity towards reactive attitudes, arguing that to the extent it is possible that an agent can change this propensity it might be better in some cases to move towards a more objective stance in our relations to other agents. I will also address the possibility of rejecting moral responsibility while maintaining some semblance of our reactive attitudes, arguing that a large portion of our social attitudes can be maintained while still rejecting the system of moral responsibility. This is due to the various types of responsibility that exist which lack the types of sanctions justified by moral responsibility. In section 2, I will argue that criminal punishment does not necessarily depend on moral responsibility. This is due to consequentialist justifications of punishment which are not directly linked to the notion of blame. I will also argue that the practice of criminal punishment is not as effective when it comes to social cohesion as is claimed, especially when compared to relevant non-punitive alternatives. By showing that both the reactive attitudes and criminal punishment are either not beneficial or not dependent on moral responsibility I hope to undermine the intuition that moral responsibility is practically necessary. In section 3, I will add some closing thoughts on this debate.

3.1 Participant Reactive Attitudes and the Moral Sentiments

The participant reactive attitudes, as described by Peter Strawson (2013), are the everyday expressions of attitudes such as resentment, gratitude, forgiveness etc., involved during our participation in the interpersonal relationships we have with other human beings. If someone were to step on my foot with the malevolent intent to cause harm against me, it seems appropriate that I react to this with some sort of indignation or a possible demand for an apology. This holds also for the more positive aspects of human relations such as giving thanks to another individual who has done you a favour. Without the expression of appreciation for the favour given, the individual might be reluctant to provide further favours in the future by holding resentment over the lack of appreciation of their giving of the favour. Third parties will also view these slights of reciprocation, or moments of appropriate gratitude, and use them as bench marks for future interactions with you. Our reactive attitudes serve as signals of our appropriate or inappropriate interactions with others which help those participants in the human community function accordingly with the knowledge so given. It also expresses that we have demands, in so far as we expect others to be capable, on others who decide to harm us or on those who have been the recipients of our acts of good will. Their failure or success at showing the appropriate response to our reactive attitudes serves as a means of inclusion or exclusion as participants in the relevant human community.

The worry, expressed by Strawson, is that if we forgo moral responsibility we would have to forgo these reactive attitudes as well, and that this fact would lessen the richness of our social interactions, or as he puts it, we would be choosing “between our humanity and our intelligence” (p. 344). Legal scholar Stephen Morse (2013) shares the assessment that the denial of moral responsibility is problematic when he makes the claim that moral responsibility skepticism “bleaches the soul” (p. 131). The problem, as they see it, in choosing the side of intelligence, in Strawson's terms, by perceiving individuals as objects as opposed to active participants in the human community, is to neglect something that is inherently human and worth keeping. Sally Satel and Scott Lilienfeld (2013) add that

> ...a blameless world would be a very chilly place, inhospitable to the warming sentiments of forgiveness, redemption, and gratitude. In a milieu where no individuals are accountable for their actions, the so-called moral emotions would be unintelligible. If we no longer brand certain actions as blameworthy and punish transgressors in proportion to their crimes, we forgo precious opportunities to reaffirm the dignity of their victims and to inculcate a shared vision of a just society. By failing to reflect the moral values of the citizenry, which encompass fair punishment, the law would lose some, if not most, of its authority. (p. 146-147)

Those who argue against these Strawsonian positions claim that even though some of our reactive attitudes would need to change, most of our everyday normal interpersonal relations will remain the same (Waller, 2011; Pereboom, 2001). The question I will first turn to is whether or not change, with regards to abandoning some of our less desired reactive attitudes is possible, and if it is, to what extent? This case will be made by focusing on three areas of evidence regarding change, historical change, experimental evidence for attitude change, as well as the legal change regarding those with mental disabilities. In the process I will elaborate on some reasons why change might be desirable if it is in fact possible. Such that some of our reactive attitudes, including but not limited to resentment, are unfair and emotionally problematic for the well-functioning of society.

3.1.1 The Argument from Impossibility

Several theorists, including Strawson, argue for, or strongly imply, that it would be impossible to remove the reactive attitudes altogether. When addressing the hard determinist position, Strawson argues

> Finally, to the further question whether it would not be *rational*, given a general theoretical conviction of the truth of determinism, so to change our world that in it all these attitudes were wholly suspended, I must answer, as before, that one who presses this question has wholly failed to grasp the import of the preceding answer, the nature of the human commitment that is here involved: it is *useless* to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do. (p. 348)

Similarly, Dennett argues that

> The utopian idea that we might reform human nature so much that punishment was no longer practically necessary is about as realistic as the idea that we could reform human biology so much that food and water would no longer be a 'practical' necessity for human life. (2012b)

There is a theoretical possibility that they are correct in the sense that ultimately changing human nature is for all practical purposes impossible. However, not all of our attitudes are such that they are impossible to change. We can train our reactions to everyday interpersonal relationships such that they can lessen their force upon us. In saying that the reactive attitudes are innate, in the sense that they are impossible to change, these theorists are admitting that we lack control over this aspect of what constitutes our behaviour. However, there are moments where they admit some change is possible, as when Dennett (2003) explains "We now uncontroversially exculpate or mitigate in many cases that our ancestors would have dealt with much more harshly" (p. 290). If by 'impossible' they mean to say we cannot change, then the system of moral responsibility is in some sense a fact about the world which cannot be justified beyond merely being the result of unchangeable human interactions. This fatalistic assessment of the nature of reactive attitudes seems to save moral responsibility at the expense of our morally-relevant capabilities, which are supposed to make moral responsibility worth wanting in the first place. The system would not be set up to reflect our skill in being better moral arbiters, but is simply there because human beings cannot help but react in ways which reflect resentment, indignation, praise and other attitudes which reflect our moral nature. If, however, we can change the extent to which we can moderate our reactive attitudes, then it is possible that we can minimize their negative effects while maintaining aspects which are more conducive to a better functioning society.

In many ways there is historical precedence for these types of changes. In the not-too-distant past many North Americans openly admitted to being disgusted by or felt superior to African Americans as well as homosexuals. These sentiments partly served as the justification for the exclusion of such individuals from civic life. There is evidence that these sentiments and attitudes have faded over time – although not completely – which serves as a reason to believe that such change is possible. These reactions were irrational manifestations of false ideas with regards to these individuals. If, by extension, we are capable of perceiving our participant reactive attitudes, which are manifest in the emotions of resentment and indignation, as misplaced and possibly false, then it is possible, by comparison to the change of attitudes and negative reactions towards African Americans and homosexuals that we can begin to diminish their effect on our normal interpersonal relationships.

Satel and Lilienfeld see this kind of change, regarding our attitudes towards racial minorities and homosexuality, as possible, but claim that these changes were due to our sense of fairness and justice which are human cultural norms shared to a high degree by cultures all over the world (p. 146). Given how deeply rooted the sense of fairness and justice is, and how prevalent it is within varying cultures, they see this higher level of change, change to justice and fairness itself, as being highly unlikely. Changes in behaviour towards other cultural groups are really a reflection of appropriately calibrating our reactive attitudes to those who were previously mistreated. I agree with their position, in that I find fairness and justice to be important aspects of human relations which would be close to impossible to eliminate, if that would even be desirable. However, it is because of our sense of fairness that I think their position ultimately fails. Along with fairness there exists competing norms, or norms which are in some sense co-extensive with fairness. Justice in a retributive sense contains the notion of desert, but in order to deserve a particular reaction that reaction and the justification for deservedness must be fair. If some of our reactive attitudes are unfair, then they should be changed or mitigated. The moral responsibility skeptic's goal is not the entire elimination of the moral sentiments, including these higher level norms, but the elimination, or reduction, of those sentiments which are not justified by our sense of fairness. It is true that the attitudes held towards African Americans or homosexuals were not replaced by completely objective attitudes, but the superficial features which served to justify these reactions were no longer seen as relevant features for these particular reactions, let alone relevant for any reactions at all. It is in this sense that attitudes became *more* objective rather than *completely* objective.

Waller (2011) argues that the fairness intuition mitigates various retributive responses such as striking back when harmed (p. 312). Waller reflects on the nature of "moral dumbfounding", which is our inability to provide reasons for our moral decisions, and suggests that change might be difficult but not impossible. Although we have deep intuitions regarding notions of justice, such as revenge and fairness in the sense of an 'eye for an eye', we have slowly been able to use our ability to reason to develop more humane systems which attempt to ameliorate these kinds of human dispositions, without necessarily removing them completely. After all, our current justice system tries to implicate justice as being blind. The ideal of justice being blind is to lessen the extent that our personal feelings of revenge and retribution impact court decisions. This change in emotionally reactivity is in an attempt to be fairer in our judicial proceedings. It is often our goal to make the legal system less influenced by our immediate heightened and agitated emotional states and responses and instead prefer the types of reasoning available at times when we are more cool and collected. This means that the ideal of impartiality is in conflict with the reactive attitudes which are justified by moral responsibility. Dennett (2003) sees this goal as foolhardy, at least in the extreme. He calls this "Kantian ideal" a fantasy "...which you somehow strengthen your pure-reasoning muscle to such a fine pitch that you can make pure, emotionless judgments untainted by tawdry guilt feeling or base longings for love and acceptance" (p. 213). I think

Dennett is right that pure emotionless judgments are not possible, but some attitudes such as resentment and indignation tend to cause more harm than they help. This does not mean emotion is not important, but that emotions, especially at times when they can make it hard to concentrate, can be an impediment to thinking.

Waller (2011) argues that "Because deeper understanding is on the side of rejecting moral responsibility, there is a significant opening for genuine deliberation to push us out of our moral responsibility myopia and toward the acceptance of an alternative perspective" (p. 314). This deeper understanding is often reflected in the growing knowledge of our cognitive faculties, which make salient the various ways in which agents can differ in ability. Satel and Lilienfeld find it unlikely that this deeper understanding referred to by Waller, such as the advances in knowledge provided by cognitive neuroscience, will dislodge praise and blame. But it is precisely the unfairness inherent in our differing cognitive abilities which these advances in knowledge are bringing to light. However, it is a separate question whether being given information regarding our cognitive processes would be enough to change people's behaviours, and whether this change of behaviour is something which ultimately improves social welfare.

To address this first point, there is evidence that being given neuroscientific information lessens the extent to which individuals praise or blame others. Nahmias, Coates and Kvaran (2007) exposed one group of participants in their study to psychological explanations of behaviour and another group to neuroscientific explanations of behaviour. They found that only 41% of the participants exposed to the neuroscientific explanation thought that an agent is morally responsible compared to the 89% in the psychological explanation condition. In other words, participants are much less likely to see an agent as responsible when their behaviour is explained in neuroscientific terms, such as chemical reactions and neuronal processes, than when it is explained in psychological terms, such as beliefs, desires, or plans. A possible explanation of these results, offered by Nadelhoffer et al. (2013), is that neuroscientific explanations bypass our folk psychological beliefs (p. 198). This means that with the presence of psychological explanations, our intuitions about the effectiveness of beliefs, desires, or plans are salient enough to still maintain moral responsibility and that somehow the neuroscientific explanations make these features less salient. Another possibility that I think is more likely, is that when beliefs, desires and plans become more salient, individuals are more likely to attribute the behaviour to the individual rather than to constitutive facts which factor into the agent's development. By putting things into the context of neuroscience, it becomes more salient that we – whatever this 'we' is – are less in control of certain factors which play a large role in our development and the relevant abilities that are manifest therein. This as I explained in the previous chapter is an affront to our sense of fairness and could explain why participants who judged the neuroscientific descriptions are more motivated to lessen their degree of moral responsibility.

Whether this is the best account available for these results remains to be seen, but what matters for my purpose here is that new information does indeed change attributions of responsibility, whether or not it is changed in a way which is more favourable to my position. This means it is possible that this change in responsibility attributions can indeed affect our behaviour towards others, as the lowering of a desire to hold people morally responsible attests. As Joshua Greene and Jonathan Cohen (2004) explain "This change in moral outlook will result not from the discovery of crucial new facts or clever new arguments, but from a new appreciation of old arguments, bolstered by vivid new illustrations provided by cognitive neuroscience" (p. 1775). Just as society's views of homosexuals has changed given the more salient understanding of the true nature of homosexual behaviour due to their visibility in society and a better understanding of the science of sexuality, so could society's views of praise or blame be altered given the more robust picture of human cognition and behaviour offered by the field of cognitive neuroscience. These kinds of changes are already being reflected in the worlds varying legal systems as they have been shifting away from blaming those who are deemed not capable of understanding the true nature of the criminal act that is perpetrated since roughly the 19th century. This shift in legal attitudes coincides with the development of psychology as a legitimate scientific practice. The existence of the legal judgment of 'not criminally responsible' on the basis of mental disorder is in some sense a concession to the view that change is in fact possible, although it is a separate question whether or not this is desirable.

When considering the upbringing of an individual who turns out to do something atrocious, we tend to find situational factors, or cognitive features, which help explain how the agent became the person who committed the atrocious act. As Pereboom (2001) explains

> Upon absorbing such information, not everyone relinquishes his attitude of indignation completely, but this attitude is at least typically tempered. It is not only that we are persuaded to feel pity for the criminal. Not implausibly, our attitude of indignation is mitigated by our coming to believe that there were factors beyond his control that causally determined certain aspects of his character to be as they were. (p. 95-96)

I would add that this lack of control, manifest in our lack of ability to impact the constitutive luck which factors into what shapes our differing abilities, is ultimately what serves as the mitigating force, given its affront to fairness.

This type of position is often criticized for medicalizing all of human behaviour (Morse, 2013), given that we are applying the same standards which mitigate blame towards, for example psychopaths, to all human beings. To some extent I think this is the inevitable outcome of a better understanding of human behaviour and cognitive science, since it exposes the extent to which we all have differing cognitive abilities. After all, what differentiates otherwise normally functioning human beings from psychopaths is in some sense a matter of degree. Psychopaths are not utterly incapable of having morally-relevant control, but their capabilities are, in some sense, severely diminished or altered in comparison to others. The differences in ability between most agents might be smaller than the gap between what we tend to consider ‘normally functioning’ human beings and those who suffer or are impacted by cognitive ‘abnormalities’, but no matter how small these differences are they still matter. The more precise our ability to assess these differences the more we can see why it would be unfair to treat as equals those who cognitively differ in ability. Whether this medicalization of all human behaviour can be conceived of in ways which help to benefit social welfare or not will be taken up in the last section, for now it is enough to address to what extent the reactive attitudes become mitigated when we adopt this stance towards those individuals society deems incapable of standard human relations.

Strawson describes the result of this kind of objective stance as follows

> To adopt the objective attitude to another human being to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided... But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. (p. 344)

He argues that this type of stance can be reasonably adopted towards individuals who for lack of ability cannot be adequate participants in normal human relations. It would, however, be "practically inconceivable" for us to forgo these reactive attitudes entirely (p. 345). This is echoed by Gary Watson (2011), who argues that psychopaths lack the ability to engage in normal human relations, specifically in ways which respond to the demands of society. According to Watson, it makes no sense to demand something of someone who is not capable of recognizing or responding to that demand (p. 314). Implicit in these positions is that only certain individuals should be dealt with objectively, but I would argue that in suggesting that an objective stance is possible for some individuals, Watson and Strawson expose the extent to which it is in fact possible to change our reactive stance towards individuals in general. I see no reason to think this type of stance could not be expanded to other individuals, let alone be inculcated in a change of moral behaviour for most individuals. In other words, if we can change our behaviour to a subset of the population, we can, if justified, change our behaviour towards other members of the human community as well.

Contrary to the position that our reactive attitudes are robustly incapable of being altered, there is plenty of evidence to suggest change is possible. It might be the case that change, though possible, might not be fully realized in that the vestiges of our evolutionary heritage might continue to be something we have to actively restrain. However, there are plenty of impulses and behaviours that most human beings find difficult to refrain from, but manage to overcome due to superseding reasons such as fairness. This difficulty should not be seen as a reason to reject the ideal of change, but could be used to highlight the areas in which we need to work harder to achieve it. Given that change is possible, we are not stuck with the unfair, emotionally agitated reactive attitudes which are justified by moral responsibility through notions of desert. Not only are these attitudes not beneficial, they are also not necessary. However, I now turn towards attitudes which are, arguably, beneficial, but I will argue these do not necessarily depend on there being moral responsibility to justify their use.

3.1.2 Maintaining Reactive Attitudes

Implicit in the previous section was that Strawson, and those who worry about the abolishment of the reactive attitudes, think that the attitudes themselves are worthwhile. Their position is that even if it were possible to remove some of our reactive attitudes, it would not be desirable. Part of this worry stems from the idea that moral responsibility skeptics want to abolish all of our reactive attitudes, removing what makes life emotionally rich and depriving it of value. I have already presented some reasons for why moving towards a more objective stance is beneficial in some instances, but this does not mean the objective stance is beneficial in all circumstances. I will now argue that even if we reject moral responsibility, this would not necessarily undermine all of the reactive attitudes that are essential for interpersonal relations. In particular, the human attitude of forgiveness can and should be maintained in a world without moral responsibility. If I am correct then there is no reason to think that the denial of moral responsibility removes the humanity from our interpersonal relationships. The denial of moral responsibility would still make life worth living. In order to make this case I will argue that some reactive attitudes, such as forgiveness, do not depend on moral responsibility. This will involve a more nuanced discussion about the various ways an individual can be said to be responsible without being morally responsible.

Part of the process involved in how these attitudes function, is the recognition of benefits and harms perpetrated by others and ourselves. Even though the system of moral responsibility is unfair, that does not mean we are incapable, in the ultimate sense, of seeing and understanding when a good deed is done or when a moral harm has been incurred. Waller (2011) puts this succinctly

> This universal denial of moral responsibility is not based on denial of rationality; rather, the claim is that there is no moral responsibility because whatever our talents and flaws, our virtues and vices, they are ultimately a matter of our good or bad fortune: ultimately, they are the product of forces we did not control. (p. 191)

As an agent who is prone to error, I can recognize in myself, often sometime after the error has occurred or when more information is provided, that I have indeed erred and caused harm to another human being. This does not mean that the harm is my fault, in a moral responsibility sense, but the harm emanates from me, and my acknowledgment of that error through a sincere apology can and often does make the situation better for all who are involved. This is because it signals to others the sincere acknowledgment of the wrongful act and the possible change of behaviour this will entail in the future.

Gratitude has similar effects, by acknowledging those who do good acts. By saying ‘thank you’ we inform the agent that we appreciate what they have done, which signals to them that other people are aware of the good things they have done. The agent does not deserve gratitude in the morally responsible desert-based sense, and sometimes demanding gratitude diminishes its worth or at the very least adds nothing to its emotional power. As quoted earlier, Strawson thinks love is also essential to moral responsibility and thinks we would lose something integral to our loving relationships without it being justified by moral responsibility, but love is often unconditional and not offered because the agent in some sense deserves it. The love of a parent towards their children is often of this kind, and it is rarely described in terms of deservedness, but instead emanates from a strong desire (Pereboom, 2014, p. 190; Waller, 2011, p. 201). Comparatively, romantic love is also provided by strong desires, and it is not clear whether desert, by deciding that someone deserves our love, is the appropriate means of achieving this emotionally rich and wonderful phenomenon. As Pereboom (2014) explains

> A decision to love on the part of another might greatly enhance one's personal life, but it is not obvious what value the decision's being free in this sense would add. Moreover, while in circumstances of these kinds we might desire that someone else make a decision to love, we would typically prefer the situation in which the love was not mediated by a decision. (p. 191)

Forgiveness, gratitude and love, which uncontroversially make life worth living, can all be maintained without the need for them to be deserved in the sense required by moral responsibility.

There is another type of responsibility at play here, which differs in kind from the type of moral responsibility argued for by those who think forgiveness, gratitude or love would not be possible given the denial of moral responsibility. Waller (2011) describes this type of responsibility as causal responsibility[6], such that I am the cause of something without being blameworthy (p. 195). A non-controversial example of this kind of causal responsibility would be if an individual opened a door, oblivious to another individual standing behind it and broke their nose. This is non-controversial since even those who advocate moral responsibility would likely think the agent should not be held morally responsible. This is because most of the features in virtue of which we hold agents blameworthy are missing from this case, such as the intent to harm the individual or some form of criminal negligence on the part of the door opener. However, even in the absence of moral blame, we can offer sincere apologies for being at the center of the causal chain which

[6] Fischer and Ravizza (1998) describe the same kind of concept, and they also call it causal responsibility (p. 1).

led to their nose being broken, and people often do apologize for these types of accidents. The apology is often issued, in this kind of case, to make the victim aware that the agent cares and is also aware of the harm that was incurred. We can even apologize for things which were initially intended by us, in that we can reflect on our past behaviour and come to see why our actions harmed others, and why that harm was not called for. The harm was not called for, in the sense that we now have new information or new beliefs, that upon reflection had we known then what we know now we probably wouldn't have done what we did.

A similar notion of responsibility is also offered by David Shoemaker (2011) called attributability responsibility, which differs from what he describes as answerability and accountability. Accountability is a desert-based form of responsibility that is comparable to what I have been describing as moral responsibility. Answerability is the degree to which agents are responsible for explaining what they have done. Attributability, on the other hand, is the extent to which we can attribute an action or result to the agent who perpetrated that action or result. For example, my cat Simone was responsible for eating the plant, in that the action is attributable to her and no one else. Shoemaker's account reflects the extent to which we should demand evaluative responses from others who are not capable. As he explains "Your demand to me to justify an attitude reflecting a groundless emotional commitment will be without a point as a demand, for I am simply devoid of the resources necessary to engage with your communicative attempt" (p. 611). This means that some agents are attributability responsible, but they are not necessarily answerability responsible or accountability responsible for their actions. My cat, for obvious reasons – the lack of comprehensive communicative skills for one – cannot answer for the actions attributable to her. However, given this, it is arguably appropriate to hold reactive attitudes towards others who are not accountable or answerable, such as a desire to avoid that individual, or to signal to them we are upset. This is because reacting in a way which signals that an individual is manipulative is not necessarily an act of condemnation; it is merely accurately attributing a character trait to that agent without making a moral demand (p. 617). As Pereboom (2014) explains, when we receive the type of evidence required to attribute to the agent a character fault, you might "take your relationship with him to be impaired to a degree that it justifies as appropriate" which is not objectionable to the moral responsibility skeptic (p. 131). This stance allows us to abandon moral responsibility in the sense of praise and blame without eschewing all of our reactive attitudes. Or as Levy (2011) explains “we can say, simply, that such an agent is *bad*, leaving open the question of whether they are also *blameworthy*” (p. 205).

It is a separate question whether a moral responsibility skeptic can hold agents answerable for their behaviour. In that the agent is capable of explaining their reasons for their behaviour upon being asked. Hearing the reasons for your behaviour, Shoemaker explains

> To the extent that these reasons fail to live up to the actual justifying reasons there are in some domain, the agent will be subject to modifications of attitudes or dispositions by various assessors. My recognition of what reasons you thought justified your action may give me serious pause, for example, about whether or not to deal with you in various ways in the future. (p. 631)

However, not all answers, if justified, deserve moral blame. For example, if I want you to throw me a surprise birthday party, and you fail to do so, your ability to answer for your behaviour does not mean that you should be sanctioned for my hurt feelings, even though I might feel differently about you, such that maybe you aren’t the friend I thought you were.

Shoemaker thinks that these types of aretaic appraisals, “judgements about the morally relevant aspects of an agent’s character in light of the agent’s attitudes or actions”, are relevant for moral responsibility (p. 612-613). These aretaic appraisals are given to individuals who are attributability responsible, answerability responsible or both. However, I think that attributability and answerability are separate from the moral responsibility required to justify the sanctions of being accountable under Shoemaker’s characterization of responsibility. Although an agent can provide reasons for their behaviour, which is attributed to them, and those reasons can affect the relationships we have with that agent, by revealing their character traits, something else is required to then hold these agents to account for their behaviour in the sense demanded by moral responsibility. To hold someone to account according to Shoemaker is “to sanction that person, whether it be via the expression of a reactive attitude, public shaming, or something more psychologically or physically damaging” (p. 623). However, lumping the reactive attitudes alongside more severe punishments is missing an important distinction I think exists when we consider agents as morally responsible. Reactive attitudes are useful in the sense that they express the various character traits which help to signify, for those capable of picking up the cues, the information necessary to more adeptly navigate social situations. But this description says nothing about the kind of blame and praise required to heap reward upon those who are lucky and physically harm those who aren't. Reward and punishment require a further justification, through the quality of deservedness offered by a system of moral responsibility, which I argue is unfair. Avoiding individuals who by their actions are not trustworthy is not unfair, in the same sense that dodging an incoming train is not unfair to the train. Unlike the train, we have rich emotional lives, but our desire to avoid harm in this way is enough to disregard the potential harm accrued by socially isolating the individual who is not trustworthy. The moral responsibility skeptic might be more prone to forgive the untrustworthy individual whose inability to maintain trusting relationships is no fault of their own, but the moral responsibility skeptic is under no obligation to put themselves in harm’s way. The untrustworthy individual does not deserve to be avoided, but their actions make them avoided nonetheless. Not out of a sense of retribution, but out of a sense of self protection.

By piling together the reactive attitudes with the harms and benefits offered by reward and punishment, the moral responsibility advocate can continue to claim that if we give up the one we have to give up the other, but I see no reason that this follows. Another possibility is that Shoemaker is conflating what Waller (2011) differentiates as character-fault and blame-fault (p. 192). Character-fault is the faults we can perceive in ourselves and other people given that we are not perfect agents, whereas blame-fault is the type of fault expressed in moral condemnation. We can be sorry for our faults, and forgive others for theirs, without the need to blame them for those faults, or hold them accountable for them. Levy (2011) thinks that this kind of discrepancy is due to the different goals of what he describes as ‘quality of will’ theorists (p. 208-210). The difference, he thinks, rests on different fields of inquiry. Quality of will theorists, such as Shoemaker, are doing work on moral psychology, which as Levy explains is only “tangentially involved” with the moral responsibility debate. As Levy makes clear

> ...if by ‘blame’ quality of will theorists mean ‘a belieflike attitude that is justified when there was ill will and a wrong act’, then they are quite right about the conditions of ‘blame’. They are quite wrong, however, in thinking that that is how the term is generally used, within the debate over free will and moral responsibility. (p. 209)

I think this is the appropriate way of reading what Shoemaker means by ‘blame’, without reading too far into what this means for moral responsibility. In conclusion, I do find Shoemaker's distinction between attributability, answerability and accountability useful, even though I think he has improperly distinguished between the justification necessary for our reactive attitudes and the justification required for moral sanctions and desert.

It is functionally possible then to maintain certain types of attitudes such as forgiveness and gratitude while denying the existence of moral responsibility. Even attitudes like resentment, which some moral responsibility skeptics find undesirable and capable of being abandoned (Pereboom, 2001, p. 200), are still useful in some situations, such as general resentment towards societal inequalities such as income disparity (Norman, 2002). However, resentment can manifest itself in ways which harm society and in such cases it might be necessary to mitigate it. Pereboom (2014) argues that resentment is particularly prone to errors “Evidence for this includes the fact that when someone blames in a way that that expresses resentment or indignation we are typically on guard that he will attribute to its target intentions and efficacy skewed in a way that would serve to justify the blame [Sic]” (p. 151). Given that change in these error prone cases is possible, and that holding reactive attitudes in some cases is justified even when moral responsibility is denied, the claim that the denial of moral responsibility is practically necessary for the reactive attitudes lacks a strong foundation. This is because the benefits derived from the reactive attitudes do not depend on moral responsibility, and the reactive attitudes that do depend on moral responsibility are arguably not beneficial at all. Therefore, the denial of moral responsibility does not mean that life would not be worth living, especially with regard to the reactive attitudes.

3.2 Criminal Punishment

Related to the concern that moral responsibility is practically necessary for the reactive attitudes, which give life an emotional richness, is that moral responsibility is practically necessary for criminal punishment. Criminal punishment, so the argument goes, is essential for the well-functioning of society and requires moral responsibility for its justification. Given that the two ideas are related my approach in many ways will be similar to the previous section. In order to undermine the claim that moral responsibility is practically necessary, I will first address the extent to which criminal punishment depends on moral responsibility. I will then look at whether or not there are indeed benefits to criminal punishment, such that maintaining the practice, in the morally responsible sense, would be practically necessary. I will argue that it is possible to justify criminal punishment without moral responsibility through a consequentialist form of justification. However, even if a retributivist or a criminal punishment abolitionist does not accept this argument, I further claim that there are relevant non-punitive alternatives which are more beneficial than criminal punishment. Given the possibility of these arguably better alternatives, the claim that criminal punishment is practically necessary is less tenable. These non-punitive alternatives are the combined practices of victim restitution, rehabilitation and quarantine.

There are two things I need to address before going forward with the argument. The first involves the extent to which this section focuses on punishment and not reward. Given that moral responsibility justifies both punishment and reward, it seems odd that this does not receive equal treatment. Part of this is due to the fact that the literature tends to focus on the punitive aspects of moral responsibility as opposed to those aspects which offer benefits to individuals. This, I think, is in part due to the asymmetrical quality of praise and blame. As Levy (2011) expresses it “...I feel keenly the unfairness of blaming wrongdoers for actions over which they failed to exercise (relevant) control. That some might be the beneficiaries of undeserved praise worries me much less” (p. 204). I share this intuition, which seems to be reflective of the psychological phenomenon of loss aversion. People tend to prefer avoiding losses than acquiring gains (Kahneman, 2011). Whether someone gains a societal reward feels less problematic than an individual losing liberty and suffering pain through the practice of punishment as sanctioned through desert-based moral responsibility. This could also be a reflection of the significance condition regarding luck in the previous chapter. Given that loss aversion is a robust psychological phenomenon, harms can be more significant for an agent than rewards, and thus are more susceptible to moral responsibility undermining luck. Of course, the more those rewards begin to shape societies in negative ways, the more significant they become. For example, certain tax policies which reward the rich at the expensive of the poor are good examples of reward based policies which are problematic, but to be fair, are probably considered problematic for individuals on both sides of the moral responsibility debate. However, if the feeling of asymmetry is not justified, it is important to note that most of what I say here can be applied to reward based policies which unduly heap benefits upon the lucky. If it is problematic to punish an individual who fails, of no fault of their own, then it is problematic to reward those who succeed, of no fault of their own. Although, given the common pronouncement of the necessity of punishment, I will proceed to address this criticism against the moral responsibility skeptic’s account.

Secondly, I think it is important to briefly describe what I mean by punishment, given that it is a loaded term with several meanings. I will be using the definition provided by David Boonin that punishment is authorized reprobative retributive intentional harm (2008). Taking these in turn, this means that punishment must be authorized by a relevant authority such as the state. It must be done to agents in a way to express disapproval of their behaviour. The agent must deserve the punishment for their morally relevant actions or inactions. The punishment must be intended, in that it can’t be a mere by-product. Finally, punishment must harm the agent, in that it leaves the agent worse off than before the punishment. Many aspects of this definition are not controversial, and to the extent that they are they will not be relevant to my argument. The only exception to this is the retributive component of punishment. It might seem that this component is rigging the game by already deciding between retributivist and consequentialist justifications for punishment, but this would be confusing what Boonin calls the nature of punishment with its justification (p. 20). In other words, if something is to be punishment, it must be performed because the agent deserved it, but whether this is done for forward or backward looking reasons is a separate question.

There is also a distinction that needs to be made between criminal punishment and our somewhat more colloquial usage of the term ‘punishment’. Given my preference for Boonin’s definition, ‘punishment’ is necessarily ‘criminal punishment’ since it is sanctioned by the state. Criminal punishment, in this sense, is a more specific form of socially mandated blame compared to the more general form of inter-personal blame discussed in the previous section. Although blame in both the specific and general sense are related it is important to highlight that they are not the same. With this distinction in mind, I will now proceed to argue that criminal punishment does not necessarily depend on moral responsibility.

3.2.1 Does Criminal Punishment Depend on Moral Responsibility?

Part of addressing this question involves distinguishing between the two justifications of criminal punishment; retributivist and consequentialist. Retributivist justifications for criminal punishment are inherently backward looking. This means that criminal punishment depends on the extent to which the agent deserves being punished in virtue of their past actions. Consequentialist justifications for criminal punishment are inherently forward looking. This means that criminal punishment is meted out because in doing so it obtains a future benefit. In some sense, the argument for practical necessity requires that criminal punishment be forward looking to some degree. This is because arguing for the practical necessity of criminal punishment is making an argument about the benefits derived from criminal punishment. If criminal punishment were to be retributively justified, then it would be justified regardless of whether the outcome would be beneficial or not. However, an argument could be made that criminal punishment is practically necessary, but only if it is meted out retributively. My argument will begin by defending the position that some forms of criminal punishment, justified for consequentialist reasons, can be maintained without requiring moral responsibility. I will then address retributivist concerns, such as the harming the innocent objection against consequentialist justifications of criminal punishment. If these objections succeed at undermining the consequentialist justification for criminal punishment, which I acknowledge is a possibility, the next section will argue that there are relevant non-punitive alternatives which are more beneficial than their punitive counterparts. This means that if criminal punishment is indeed necessary, then there are reasons to think it can be justified based on a consequentialist account without moral responsibility. But if this argument fails, there are reasons to suspect that criminal punishment is not practically necessary in the first place. Either way moral responsibility is not necessary.

Whether a consequentialist justification for criminal punishment requires moral responsibility is going to depend on the extent to which the retributive component of criminal punishment undermines the consequentialist justification. As mentioned before, this component is not related to whether or not criminal punishment is ultimately forward or backward looking. If a police officer tickets a pedestrian, after the officer witnesses a car going over the speed limit, it is reasonable to ask why the pedestrian was ticketed and not the driver of the car. If we say that the driver deserves the ticket, and not the pedestrian, it needs to be determined whether this deservedness is dependent upon moral responsibility. A consequentialist has a resource open to them to deny the conclusion that this deservedness requires moral responsibility; one which I provided for them in the previous section. The sense of ‘deserved’ here can simply refer to the causal link between the agent and the action, such that they are attributability responsible. If criminal punishment is indeed necessary, then it is possible that we can justify that punishment based on the future benefits of doing so, and on the extent to which we can attribute the criminal actions to the agent, while at the same time denying moral responsibility. This kind of justification has been defended by Greene and Cohen. After addressing the views of hard determinism, libertarianism and compatibilism, they conclude that

> The forward-looking-consequentialist approach to punishment works with all three responses to the problem of free will, including hard determinism. This is because consequentialists

> are not concerned with whether anyone is really innocent or guilty in some ultimate sense that might depend on people's having free will, but only with the likely effects of punishment. (p. 1777)

Similar views have been endorsed by other cognitive scientists and philosophers, such as Michael Gazzaniga (2011, p. 213), Paul Thagard (2010, p. 205-206), Patricia Churchland (2011, p. 81), Steven Pinker (2011, p. 538-540), as well as many others. Although this is a popular solution, there are some reasonable criticisms provided by retributivists.

Morse (2013), who is a retributivist when it comes to the justification of criminal punishment, is particularly put off by this idea. This is because he thinks the cost and benefit analysis of a consequentialist justification for criminal punishment can be used to justify a whole manner of practices which would be undesirable (p. 130). Morse expresses this worry in a rhetorical question "Why not have a regime of 'screen and intervene' if we are convinced that we could increase social safety?" (p. 130). This is reminiscent of the final tale in Isaac Asimov's (2004) book I, *Robot*, in which the robots decide that the only way to protect humanity is to take control over it. If humans are prone to getting hurt while driving, why not take away their cars? After all, by doing so, rates of injury and death will go down, which, at first glance, seems like a net benefit. Given that the majority of crime is committed by men aged 16-30, shouldn't we just lock them all up? Since these questions are a matter of consequentialist cost benefit analysis this means they remain a live option under a system of consequentialist criminal punishment. According to Morse, only retributivist forms of criminal punishment can prevent these undesirable outcomes by tightly linking criminal punishment to the retributive deservingness of the agent. As he puts it "it is the compatibilist view of agency, coupled with notions of desert and retribution, that is the foundation of proportionality and the protection of human rights within our desert-disease system of social control" (p. 130).

This worry is related to a common objection against consequentialist forms of criminal punishment, which is the punishing or harming the innocent objection. As Boonin explains

> If punishing Larry for robbing a liquor store is morally justified because it produces more overall utility than would any available alternative, then in at least some circumstances, deliberately punishing an innocent person is also morally justified because it produces more overall utility than would any available alternative. (p. 41)

One response to this is to argue that given the definition of criminal punishment, punishing innocents is by definition not possible. Although this is unsatisfactory since even if the agents harmed are not technically punished, their harm is still undesirable regardless of what we call it. This leaves it to the consequentialist to either accept or deny this consequence; the former consequence being unpalatable while the later consequence seems unlikely.

I lack a satisfactory response to these concerns, partly because, for reasons which will be clear in the next section, I find there to be relevant non-punitive forms of criminal justice which would render criminal punishment completely unnecessary. It could be the case that Morse’s worry and the harming the innocent objection are not that concerning, such that biting the bullet and accepting these consequences is worth the benefit of maintaining criminal punishment. I doubt this is the case, but not because criminal punishment depends on moral responsibility, but because I don’t think criminal punishment is necessary at all. However, if criminal punishment is necessary, I think the consequentialist can at least challenge the notion that it depends on moral responsibility. One response is to point out that retributivist justifications have similar harming the innocent implications. If criminal punishment depends on desert, then you are not supposed to punish those who do not deserve it. Given that it is impossible to have a perfect criminal justice system, such that only those who are guilty get punished, it is impossible to avoid the fact that retributive criminal punishment allows for the punishment of those who don’t deserve it (Boonin, 2008, p. 99). A consequentialist can handle this problem by accepting the punishment of innocents given the net benefits of the practice in general, while arguing that consequentialist justifications also require that criminals, even those falsely accused, be harmed less. This is because they are not stringently tied to the notion that a worse crime deserves a harsher punishment, which is a possible implication of retributivism.

Whether or not these lines of argumentation ultimately succeed is not as important for my argument. My main point is to argue that it is at least in principle possible to justify criminal punishment without the need for moral responsibility. It is therefore not obvious that criminal punishment is dependent upon moral responsibility. However, if I can convincingly show that there are better alternatives to criminal punishment, then I would have more thoroughly succeeded at my goal of rendering moral responsibility unnecessary.

3.2.2 Better Non-Punitive Alternatives

Dennett and Morse both directly ask criminal punishment abolitionists to provide alternatives (Dennett, 2012a; Dennett, 2012b; Morse, 2013, p. 131). If alternatives don’t exist, there is no sense discussing getting rid of the current practice we have that they say works. Their responses to the alternatives that do exist, are either that they find them unappealing for various reasons or think that the evidence is not yet robust enough for their wide spread adoption. Part of the difficulty is that it is hard to implement non-punitive practices in a society which perceives criminal punishment as intuitively necessary for the well-functioning of society. This means that the quality of evidence that is available is not conclusive on either side of this debate. My position here will be to argue that the evidence is at least promising and worth considering. There are three alternatives I will be advancing here which could be practiced alone, in tandem, or to various degrees as necessary. These alternatives are pure restitution, rehabilitation and quarantine. I will describe each in turn, providing, though not exhaustively, what evidence there is in favour of their adoption. I will then address whether or not their adoption would be appealing, such that there could be negative and unwanted side effects. These side effects are similar to the concerns raised by Morse against consequentialism, as well as the negative implications of ‘funishment’. If these alternatives could benefit society such that criminal punishment is no longer necessary, then the denial of moral responsibility is a viable option.

The first alternative is the idea of pure restitution as advocated by David Boonin (2008, p. 218-220), which shares some resemblance to the idea of restorative justice. The idea is that the government should compel those who perpetrate illegal acts to compensate victims for the harms incurred by their actions. In a similar vein, restorative justice requires that the perpetrator heal the societal wound they created by their actions (Gazzaniga, 2011, p. 208). What makes this not a form of criminal punishment is that the intent of such a practice is not to make the perpetrator suffer. This also does not require that the perpetrator be morally responsible, since all that is required is that the perpetrator be the cause of the harm in that it is attributable to them.

It might be objected that pure restitution is not pragmatically different from criminal punishment, such that being compelled to pay back a victim is not all that different from being criminally punished with a fine of similar value. Intuitions about this problem might be motivated based on how the individuals are compelled, and what that entails, for example, whether restitution is demanded or simply on offer. There is no reason to think that this type of system needs to be brutally enforced, although, some may think compliance would decrease if it was not. All of these problems might be an issue if pure restitution is considered in isolation, but it might simply be an easy first step in our justice system, for those who are aware that they have erred to make amends, with treatment or quarantine being used if restitution is not a viable option for whatever reason.

There is some evidence regarding the success of this form of justice. In the few areas where it has been adopted, pure restitution has been indicated in reducing the frequency of crime, maintains a high satisfaction rating from those who have participated in the practice, whether victim or perpetrator, as well as evidence that it has saved governments money (Daniels, 2013, p. 312). Similar practices have also been implicated in bringing about peace to civil and international conflicts, such as in apartheid South Africa (Pinker, 2011, p. 543-546). One negative consequence of such a practice is that it fails to adequately deal with individuals prone to more violent tendencies (Dennett, 2012a), but this also is only problematic if pure restitution is practiced in isolation.

For those who are unable to participate in such a program, due to extreme violent tendencies or a lack of regard for others etc., the next best option available, so long as it is possible, is rehabilitation in the form of treatment. The rehabilitation of criminals is harder to assess evidentially, since there are many ways to rehabilitate someone, some more efficacious than others. However, this is more likely an indication of our current state of scientific knowledge, and there is no reason to think that better and broader more efficacious treatments would be delivered in the future. One concern about the treatment model is that it seems to imply that criminals are in some sense diseased, or abnormal, and need to be made well or normal. It is possible that some criminals are indeed diseased or abnormal, but it need not be the case that they all are. Some individuals become criminals due to ignorance, or a lack of better options. Creating better societies, such as decreasing income inequality, increasing welfare programs, offering universal health care etc. should be an integral component towards a decrease in crime, and are options available for the moral responsibility skeptic. However, for those with a substantial amount of bad luck, and who do not suffer from any clearly defined clinical problems, treatment could merely be steps towards making the perpetrators lives better by removing the incentives for crime, or by educating them when they are ignorant of the harms they have caused. For example, evidence exists that correctional education reduces recidivism rates (Chappell, 2004).

For those who are criminals and do suffer from some form of mental illness, our western system of criminal justice is already beginning to incorporate various programs towards helping or ameliorating their symptoms. To this end, there are some treatments which are evidentially successful, such as decreasing the recidivism rate of individuals diagnosed with psychopathy (Polaschek, 2014). However, more successful treatments are needed, and not all treatments are as effective.

If treatment is not possible, the last option available, as advocated by Pereboom (2013), is that these individuals would be quarantined (p. 53). This type of incapacitation through quarantine does not mean that the criminal would live in squalor. This is because Pereboom (2014) endorses "an axiological moral theory which includes better consequences as valuable, where morally fundamental rights being honored and not violated count among the good consequences" (p. 172). This means that the humane treatment of those we incapacitate is demanded by our sense of fairness towards those who to no fault of their own, cannot abstain from injuring others in the community. This type of consequentialist position shares the sense that a better world is morally valuable, but has the added benefit of maintaining our fundamental sense of fairness which is undermined in a system of moral responsibility.

The more, in some sense, repulsive view for criminal punishment advocates are the combination of rehabilitation and quarantine. This is because they maintain some of the concerns raised by Morse against consequentialist justifications for criminal punishment. In particular, what is concerning is the idea of 'screen and intervene'. Pereboom (2014) responds to this worry by pointing out that even if we had the technology to 'screen and intervene', such that it would be free from error, allowing us to incapacitate those who have yet to commit a crime, that this in fact would not be an affront to human decency. He explains that "it is important to understand that the incapacitation account will specify that the circumstances of such detention would not be harsh, and that allowing the agent to be reasonably comfortable and to pursue fulfilling projects would be given high priority" (p. 171). Even if we were capable of determining within a range of probabilities whether or not an agent would be likely to commit a social harm, it is difficult to compare the sense of injustice of such incapacitation with the probabilistic benefit so delivered. It is possible that our sense of justice when it comes to being incapacitated is incommensurable with the good obtained by preventing a potential future crime, making the trade-off between these options impossible to define. When a crime has been attributed to an agent in a court of law, and it is determined that the agent is capable of future harm, heavy monitoring combined with possible quarantine could then be justified by a moral responsibility skeptic. This is due to the evidence involved, and not because the individual deserves the treatment. However, just as the current criminal justice system, which advocates criminal punishment, is imperfect, so might our doling out of treatments to people who were not attributability responsible for their criminal actions be imperfect. The added benefit of the treatment and quarantine model is that at least those who did not actually commit the crime, would not be directly harmed because of it.

Saul Smilansky (2011) attacks these accounts of criminal justice from the opposite direction. If, like Pereboom suggests, that incapacitated agents would be made “comfortable”, and that there well-being will be given “high priority”, then this seems to mean that the moral skeptic is committed to what Smilansky calls ‘funishment’. According to Smilansky

> Funishment would resemble punishment in that criminals would be incarcerated apart from lawful society; and institutions of funishment would also need to be as secure as current prisons, to prevent criminals from escaping. But here the similarity ends. For institutions of funishment would also need to be as delightful as possible. (p. 355)

This is an implication of these non-punitive accounts, according to Smilansky, because “Since [moral responsibility skepticism] holds that no one deserves the hardship of being separated from regular society, this hardship needs to be compensated for” (p. 355).

Smilansky's argument takes the form of a reductio ad absurdum, in that the soaring costs necessary to implement such a policy would be “horrendously expensive” (p. 356). Levy (2012) counters this claim by arguing that the adoption of moral responsibility skepticism would actually decrease costs financially and socially as well as morally. As he explains

> To the extent to which believing in moral responsibility is an obstacle to shortening sentences and equipping offenders with skills and opportunities for honest work (since doing so seems to reward offenders, making them better off in the long-term), we can point to the cost of *retaining* the concept of moral responsibility and ask whether we should be willing to pay it. (p. 491)

Our ability to lower recidivism rates would also make indefinite incapacitation less likely, and if we create a better society economically and increase the availability of health benefits and education, these costs will continue to decrease, making the cost of ‘funishment’ feasible compared to our current practices of criminal punishment.

Given the non-punitive alternatives ability to lower costs, decrease recidivism rates, decrease crime; all the while minimizing the harmful treatment of those who commit crimes, they remain very appealing options. While acknowledging that the evidence so far is still preliminary, I think it is at least tentatively reasonable to consider these alternatives as live options. This would mean that criminal punishment is not necessary, and given that the non-punitive alternatives do not require moral responsibility for their justification, moral responsibility itself is not necessary for criminal justice.

3.3 Conclusion

This chapter has been an attempt to motivate moral responsibility skepticism by dispelling the myth that certain aspects of our interpersonal lives, including the practical necessity of punishment, depend on moral responsibility. In light of our reactive attitudes I argued that we can maintain some of them while getting rid of others all the while still denying moral responsibility. As for criminal punishment, I argued that it is possible for some consequentialist justifications of criminal punishment to not require moral responsibility, but added, if this wasn’t convincing enough, that there remain non-punitive alternatives which are better than our current practices of criminal punishment. If this argument was successful, then moral responsibility is not practically necessary, either for the reactive attitudes worth keeping, or for the beneficial aspects of criminal justice. This means that moral responsibility skepticism remains a viable option.

Conclusion

The purpose of this thesis is to argue in favour of moral responsibility skepticism. I began by re-positioning the debate away from questions involving determinism and towards issues involving control. Offering my own account of agnostic control, which is a history-insensitive account, that is agnostic about the truth of determinism. Although, I argue that agnostic control is the best account available for those who wish to maintain moral responsibility, my position is that this ultimately fails. It fails due to the contrastive differences in ability regarding morally-relevant control. These differences are the result of constitutive luck, which render moral responsibility judgments unfair.

To the extent I succeeded at arguing for moral responsibility undermining luck, I proceed to address various criticisms that are frequently levelled against moral responsibility skeptics. The first such criticism is that moral responsibility is 'fair enough'. I argue that given the significance of the moral responsibility system, due to its distribution of rewards and punishments, we should try to do better than settling for society to be 'fair enough'. The next two criticisms focus on the extent to which moral responsibility is practically necessary, such that without it, life would be devoid of those aspects which give it meaning, and that our institutions which serve the well-functioning of society would no longer be justified. I argue that life would continue to be meaningful with the denial of moral responsibility. This is because some of our reactive attitudes, justified by moral responsibility, should be mitigated, such as various aspects of resentment. However, other reactive attitudes such as forgiveness, love and gratitude could, and should, be maintained even when moral responsibility is denied. Lastly, I argue that punishment is not necessarily dependent upon moral responsibility, but even if it were, there are non-punitive alternatives which are possibly better and do not require moral responsibility for their justification.

This is not an exhaustive assessment of the various criticisms against moral responsibility skepticism, but contrary to the majority opinion against this view, I hope to have presented moral responsibility skepticism as a viable option that should at least be taken seriously.

Bibliography

Asimov, S. (2004). I, *Robot*. New York, NY: Bantam Dell.

Blackmore, S. (2012). *Consciousness: An Intorduction (2nd ed.).* New York, NY: Oxford University Press.

Boonin, D. (2008). *The Problem of Punishment.* New York, NY: Cambridge University Press.

Chappell, C. A. (2004). Post-Secondary Correctional Education and Recidivism: A Meta-Analysis of Research Conducted 1990-1999. *Journal of Correctional Education,* 55(2), 148-169.

Churchland, P. (2011). *Braintrust.* Princeton, NJ: Princeton University Press.

Clark, T. W. (2012, October). Clark Comments on Dennett's Review of *Against Moral Responsibility* [Review of the book *Against Moral Responsibility*]. *Naturalism.org.* Retrieved from http://www.naturalism.org/resources/book-reviews/exchange-on-wallers-against-moral-responsibility

Clarke, R. (2003). *Libertarian Theories of Free Will*. New York, NY: Oxford University Press.

Clarke, R. (2005) Agent Causation and the Problem of Luck. *Pacific Philosophical Quarterly*, 86, 408-421.

Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, Aggression, and the Southern Culture of Honor: An 'Experimental Ethnography'. *Journal of Personality and Social Psychology*, 70(5), 945-960.

Daniels, G. (2013). Restorative Justice: Changing the Paradigm. *Probation Journal*, 60(3), 302-315.

Dennett, D. C. (1984). *Elbow Room.* Cambridge, MA: MIT Press.

Dennett, D. C. (2003). *Freedom Evolves.* New York, NY: Penguin Books.

Dennett, D. C. (2012a, October). Dennett Review of *Against Moral Responsibility* [Review of the book *Against Moral Responsibility*]. *Naturalism.org.* Retrieved from http://www.naturalism.org/resources/book-reviews/dennett-review-of-against-moral-responsibility

Dennett, D. C. (2012b, October). Dennett's Rejoinder to Clark [Review of the book *Against Moral Responsibility*]. *Naturalism.org.* Retrieved from http://www.naturalism.org/resources/book-reviews/exchange-on-wallers-against-moral-responsibility

Dennett, D. C. (2014, January 24). Reflections on *Free Will* [Review of the book *Free Will*]. *Naturalism.org*. Retrieved from http://www.naturalism.org/resources/book-reviews/reflections-on-free-will

Fischer, J. M. (2012). Semicompatibilism and Its Rivals. *The Journal of Ethics,* 16, 117-143.

Fischer, J. M., & Ravizza, M. S. J. (1998). *Responsibility and Control.* New York, NY: Cambridge University Press.

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy,* 66(23), 829-839.

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

Freeman, S. (2007). *Rawls.* New York, NY: Routledge.

Gazzaniga, M. S. (2011). *Who's in Charge: Free Will and the Science of the Brain.* New York, NY: HarperCollins Publishers.

Greene, J., & Cohen, J. (2004). For the Law, Neuroscience Changes Nothing and Everything. *Philosophical Transactions of the Royal Society of London B,* 359(1451), 1775-1785.

Haidt, J. (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion.* New York, NY: Pantheon Books.

Harris, S. (2012). *Free Will.* New York, NY: Free Press.

Hood, B. (2012). *The Self Illusion.* Toronto: HarperCollins Publishers LTD.

Hume, D. (1777). Essays and Treatises on Several Subjects. *Davidhume.org.* Retrieved from http://www.davidhume.org/texts/etss.html

Kahneman, D. (2011). *Thinking, Fast and Slow.* Canada: Doubleday Canada.

Kane, R. (1999). Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism. *The Journal of Philosophy,* 96(5), 217-240.

Kane, R. (2009). Libertarianism. *Philosophical Studies,* 144, 35-44.

Levy, N. (2011). *Hard Luck.* New York, NY: Oxford University Press.

Levy, N. (2012). Skepticism and Sanction: The Benefits of Rejecting Moral Responsibility. *Law and Philosophy,* 31, 477-493.

Levy, N. (2014). *Consciousness and Moral Responsibility.* New York, NY: Oxford University Press.

McCann, H. J. (2012). Making Decisions. *Philosophical Issues,* 22, 246-263.

Mele, A. R. (1995). *Autonomous Agents.* New York, NY: Oxford University Press.

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

Mele, A. R. (2009). *Effective Intentions.* New York, NY: Oxford University Press.

Mele, A. R., & Robb, D. (1998). Rescuing Frankfurt-Style Cases. *The Philosophical Review,* 107(1), 97-112.

Morse, S. J. (2013). Compatibilist Criminal Law. In T. A. Nadelhoffer (Ed.), *The Future of Punishment* (pp. 107-131). New York, NY: Oxford University Press..

Nadelhoffer, T., Gromet, D., Goodwin, G., Nahmias, E., Sripada, C., & Sinnott-Armstrong, W. (2013). The Mind, the Brain, and the Law. In T. A. Nadelhoffer (Ed.), *The Future of Punishment* (pp. 193-211). New York, NY: Oxford University Press.

Nagel, T. (2007). Moral Luck. In L. P. Pojman (Ed.), *Ethical Theory: Classical and Contemporary Readings (5th ed.)* (pp. 294-302). Belmont, CA: Thomson Wadsworth.

Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions. *Midwest Studies in Philosophy,* 31(1), 214-242.

Norman, R. (2002). Equality, Envy, and the Sense of Injustice. *Journal of Applied Philosophy,* 19(1), 43-54.

Pereboom, D. (2001). *Living Without Free Will.* New York, NY: Cambridge University Press.

Pereboom, D. (2013). Freewill Skepticism and Criminal Punishment. In T. A. Nadelhoffer (Ed.), *The Future of Punishment* (pp. 49-78). New York, NY: Oxford University Press.

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

Pinker, S. (2011). *The Better Angels of our Nature: Why Violence has Declined.* London: Viking.

Polaschek, D. L. L. (2014). Adult Criminals With Psychopathy: Common Beliefs About Treatability and Change Have Little Empirical Support. *Current Directions in Psychological Science,* 23(4), 296-301.

Pritchard, D. (2005). *Epistemic Luck.* New York, NY: Oxford University Press.

Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The Seductive Appeal of Mindless Neuroscience.* New York, NY: Basic Books.

Shermer, M. (2001). *The Borderlands of Science: Where Sense Meets Nonsense.* New York, NY: Oxford University Press.

Shoemaker, D. (2011). Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. *Ethics*, 121, 602-632.

Smilansky, S. (2011). Hard Determinism and Punishment: A Practical Reductio. *Law and Philosophy,* 30, 353-367.

Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies,* 75, 5-24.

Strawson, P. (2013). Freedom and Resentment. In R. Shafer-Landau (Ed.), *Ethical Theory: An Anthology (2nd ed.)* (pp. 340-352). Oxford: Blackwell Publishing Ltd.

Thagrad, P. (2010). *The Brain and the Meaning of Life.* Princeton, NJ: Princeton University Press.

Waller, B. N. (2011). *Against Moral Responsibility*. Cambridge, MA: MIT Press.

Waller, B. N. (2012, October). Reply from Waller [Review of the book *Against Moral Responsibility*]. *Naturalism.org.* Retrieved from http://www.naturalism.org/resources/book-reviews/exchange-on-wallers-against-moral-responsibility

Watson, G. (2011). The Trouble with Psychopaths. In R. J. Wallace, R. Kumar, & S. Freeman (Ed.), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (pp. 307-331). Oxford: Oxford University Press.

Wegner, D. M. (2002). *The Illusion of Conscious Will.* Cambridge, MA: MIT Press.

Weinraub, B. (2004, July 7). The Young Guy of 'Family Guy'; A 30-Year-Old's Cartoon Hit Makes and Unexpected Comeback. *The New York Times.* Retrieved from http://www.nytimes.com/2004/07/07/arts/young-guy-family-guy-30-year-old-s-cartoon-hit-makes-unexpected-comeback.html

Widerker, D. (1995). Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. *The Philosophical Review*, 104(2), 247-261.

Wolf, S. (1988). Sanity and the Metaphysics of Responsibility. In F. Schoeman (Ed.), *Responsibility, Character and the Emotions: New Essays in Moral Psychology* (pp. 46-62). New York, NY: Cambridge University Press.