This short dissertation is submitted in fulfilment of the requirements for a Master of Arts in Philosophy at the School of Religion, Philosophy and Classics, College of Humanities, University of KwaZulu-Natal, Pietermaritzburg Campus

Responsible Agency

Even if you couldn't have done otherwise

A defense of Reasons Responsiveness Semicompatibilism

Matthew Mumford

209515872

November 2014

As the candidate's Supervisors, we have approved this dissertation for submission

Dr Jacek Brzozowski _____ Date: _____

Dr Heidi Matisonn _____ Date: _____

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Declaration of Originality

I, Matthew Charles Mumford, declare that:

(i) The research reported in this dissertation, except where otherwise indicated, is my original work.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced.

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: .....

Date: .....

Contents

1 Introduction

A free agent might be described as someone who exercises their will and so deserves the consequences of their actions. Free will plays what seems to be an indispensable role in the functioning of society as we do not think praise or blame should be accorded for deeds done accidentally or without their being freely willed. In order to be morally responsible for our deeds good or bad, our choices and actions must be *ours* in some meaningful way. Courts of law will only convict the accused if it can be shown that the crime was committed intentionally or 'on purpose'. From the psychological perspective it seems that motivation and the meaningfulness of social interactions like being in love or showing gratitude and resentment would all lose something important if our actions were not really our own.

A complete account of free will would depend on almost every area of philosophy including how a free mind might relate to the physical world, metaphysical concerns about what is possible and what is necessary, epistemic considerations about how we might know when we are free, how freedom relates to the law and punishment, the ethics of moral responsibility, and the philosophy of physics about the laws of nature. It is thus a difficult task to concisely define free will, but intuitively we find it obvious that whatever it is, we have it.

In the face of such surefooted societal confidence in the existence of free will, there lurks a more pervasive kind of restraint on human freedom. Metaphysical consideration of the plausibility of determinism seems to cast doubt on the very existence of free will. The basic form of deterministic arguments against free will, which I will elaborate on later, is that we cannot be free if everything we do is wholly determined by the past.

Philosophers react to the consequences of determinism in two main ways. Either they are incompatibilists about free will and determinism, or they think that the two are compatible and so are compatibilists. Incompatibilists were all initially libertarians who thought that determinism must be false on account of our incontrovertible experience of freedom. A second incompatibilist view emerged called hard determinism, which took the opposing incompatibilist view arguing that due to the truth of determinism, libertarian free will must not exist.

It is the goal of this short dissertation to defend a particular brand of compatibilism. In order to adequately motivate for this position within the labyrinth that is the free will debate, I will move fairly roughshod through accounts of incompatibilism and conventional compatibilism in

order to defend a particular brand of compatibilism called reasons responsive semicompatibilism. In such muddy waters I think it useful to be upfront that I will be motivated by two assumptions, namely that our world is physicalist and determined. We don't know the truth values for determinism or physicalism, but it seems sensible to start from assumptions that seem to be most probably true based on the evidence from the sciences. I aim to show that, given these basic assumptions, a reasons-responsive semicompatibilist account of freedom is our best bet. Below I outline my argument, the defense of which will constitute the rest of the paper.

I begin with an explanation of the concept of determinism and the apparent consequences for freedom if this theory is found to be true. A view on freedom called 'hard incompatibilism' held by Pereboom (2001) takes the high probability of the truth of determinism to show not only that freedom is ruled out but also that we could never hold anyone responsible for anything. I explain the consequence argument for incompatibilism, one of the strongest versions of the standard arguments for incompatibilism presented by Van Inwagen (1975), which shows that freedom (to have done other than what was actually done) is unlikely to exist.

Next I will consider how a libertarian about freedom might respond. I will explain indeterminism and consider how quantum mechanics might make room for human consciousness to break into the causal picture, despite an otherwise determined world, to allow for responsible agency. I argue that we can either accept physicalism and reject libertarianism at this point or accept that there is something more than the physical which is causally at play. If the latter is true then it would seem that Kane's (1998) libertarian account would be a good way to spell out this intuition. In this case I think that we could have free will of the kind that allows for responsible agents that could have done other than they actually did. However if the former is correct then we must accept that human beings are determined to have done whatever we have done and that we cannot be free in this sense.

However I do not take this to mean that we are never morally responsible agents. I will consider compatibilist accounts which accept determinism but unlike incompatibilists do not think this means that free will and determinism need be mutually exclusive. Compatibilists argue that the two constructs can coexist sensibly. However, I argue that any attempt to advance an account of compatibilism that requires freedom to have done otherwise is bound to fail due to the consequence argument.

In the search for such an account I will show that although moral responsibility and freedom are normally understood to go together, there are ways in which they can come apart. If we are morally responsible then it seems that we must have been free in some sense in order to ground this responsibility. If we are genuinely free then it also seems that we are morally responsible for our actions. In crossing the compatibilist/incompatibilist divide, arguments are often leveled at either freedom or responsibility separately. Incompatibilists may argue against moral responsibility directly and inversely compatibilists argue for moral responsibility directly and by virtue of this claim that we are free. When I refer to responsible agency I am referring to both these constructs as a package. I will refer to them separately where appropriate.

Semicompatibilism argues that, rather than trying to find some way to make determinism compatible with freedom to have done otherwise, we could rather look for an account of responsible agency that simply doesn't require this criterion of 'could have done otherwise' for freedom. This could be seen as a sidestepping of the charge from determinists because it requires that we change our 'normal' understanding of free will. Semicompatibilists disagree. They argue that in defining freedom, the criterion of 'could've done otherwise' is an assumption. In other words, it is not an accepted fact that our normal understanding of freedom is essentially that we could have done otherwise. I consider three routes to semicompatibilism, all of which aim to show that this is a faulty assumption, but focus primarily on a defense of Frankfurt's famous rejection of the 'could have done otherwise' requirement. Frankfurt (1971) shows that, although in most normal cases they do go together, there are cases where morally responsible free agents could *not* have done otherwise. Given that Frankfurt cases show our normal understanding of responsible agents doesn't necessarily require freedom to have done otherwise, we should instead try to find what this 'normal' understanding of freedom might be.

Semicompatibilists like Fisher and Ravizza (1982, 1998 & 2011) offer such an account of free will and responsibility which focusses on the actual sequence of events and so is not affected by any argument that assumes free will requires genuine alternatives. The hard work for semicompatibilists is to provide a persuasive account of how responsible agency might adequately be captured without the traditional understanding of free will as the ability to have done otherwise. Their task is to pick out a plausible mechanism, or actual sequence of what counts as free action which, at the same time doesn't require alternate possibilities. The general idea is that, in line with the Aristotelian idea that humans are naturally responsive to reasons, an agent could have done otherwise if he had been differently reasons responsive at the time. I think that Fisher and Ravizza

have managed this and I aim to defend reasons responsiveness semicompatibilism against objections from some of their critics.

2 Determinism & Determinist Incompatibilism

Determinism claims that every event is determined by preceding events. Causal determinists think the universe is a closed causal system in which matter interacts, at whatever microphysical level is fundamental, in such a way that cause and effect necessitate everything that happens. Every event is caused by preceding event(s) and no event happens without a cause. A causal definition is not the only way to conceive of determinism. Van Inwagen (1975) explains determinism by considering its effect over time without appeals to causation. He argues that if we could observe the entire state of the world at T_1 and the entire state of the world at T_2, then we would observe that T_1 in accordance with the laws of nature entails T_2.

Determinists assume there must be some set of universal physical laws that apply without exception. If the fundamental nature of our world acts as per some set of mechanical rules and if humans are just a part of the natural world, then it seems that at least our brains are caused to be as they are. Human brains are made of matter and so are part of this world. If this world is deterministic the laws of nature should apply to brains in the same way they apply to everything else in the world. If we accept that what happens in our minds depends on what happens in our brains (assuming some sort of supervenience of the mental on the physical) then it seems that our minds are entirely determined to be as they are. If our minds are entirely determined to be as they are then it seems that we could never have done other than we did, that we have no control over our actions and that this means we cannot be free.

The intuition behind how the truth of determinism would affect free will (and consequently moral responsibility) can be captured informally in many ways. These formulations of the problem of free will, although not as philosophically careful as could be, are perhaps more persuasive in their rhetoric. To give a sense of how the problem of determinism for free will can be captured so variously I give a number of possible angles that could be taken at the same central idea:

> If determinism is the case then our brains are machines, ticking along just like the rest of nature, one event caused by another (Dennett, 1984). If this is so then human beings are determined to have done as they have done so we cannot be free no matter how free our subjective experience is.

Although we feel that we tie together our reasons and desires in our own particular way, everything that we do is a consequence of our genetic propensities and environmental influences both past and present, so we are not free.

If we accept that the universe is deterministic and that there is nothing more than the physical, then whatever we might be thinking, feeling or doing must be entirely the product of preceding events, each causing the next. If we are entirely determined then we cannot be free.

We will always do whatever is determined by the insentient physical interactions that happen at micro-level - whatever the fundamental level of the physical world happens to be. Therefore we cannot be free.

If determinism is true it is impossible that an agent could have done other than she did in any number of hypothetical re-runs of precisely the same set of conditions. If it is not possible to have acted other than the way we did then free will is not possible.

Incompatibilists about free will and determinism are split into two opposing camps. In the section to follow I will explore the libertarian camp which takes these arguments to show that determinism must be false. Hard determinist philosophers such as Pereboom (2001) would agree with some or all of the above arguments and take the incompatibility of free will and determinism to show that if determinism is true this entails that there is no free will. Unlike any other position in the free will debate, determinists have it easy as theirs is essentially a negative position in accepting determinism and the consequences for free will which stem from this acceptance. They don't need to defend a positive thesis of free will. Proponents of any other view must explain their way around determinism in positing a positive account of what free will is and defending their view against any attacks that hard determinists can direct at them.

Although it is not a necessary truth that our world is deterministic, many hard determinists think that determinism is in fact true at our world. Determinists must provide some justification for taking determinism so seriously. Such support can be found in the empirical evidence coming out of our best current science. Operating from within the deterministic physical world paradigm, the scientific community seeks causal explanation for the way the world works. The assumption

that matter behaves regularly (even if the fundamental physics is accurately described by quantum indeterminacy as I explain later on) allows scientists to observe, describe and predict phenomena. The success of science both theoretically and in its practical usefulness seems to be a strong source of the intuition to accept that the fundamental laws of nature operate mechanistically with no exceptions.

Lack of evidence to the contrary is a strong reason to believe determinism is true. There have been no documented cases where the laws of nature have been shown to be false. Van Inwagen (1983) argues that if the laws of nature had been contradicted, scientists would have rewritten the laws or scrapped them altogether depending on the findings. The laws of nature are descriptive rather than prescriptive in that they describe a consistent set of universal laws which are not conditional in any way. The process of science has been amassing data, refining, falsifying and updating its theories for a considerable time and the body of knowledge supporting the regularity of the laws of nature as described by science is significant. If at any stage a demonstrable irregularity was noticed scientists would have been forced by the rules of their own method to denounce its assumption of regular causation.

We have at least some reason to believe that determinism is a plausible reality. Hard determinists take this state of affairs to suggest that there is no free will and have argued as such. There are two ways to conceive of free action that can be attacked by the determinist. The first of which is best explained with the example of the Garden of Forking Paths. An agent comes to a fork in the path and goes left after choosing left. What makes this his free choice is that given the exact set of circumstances (in a re-run of exactly the same circumstances) he could have chosen to have gone right instead and have gone right. He 'could have done otherwise' is thus a criterion for free action. If other possibilities are open to the agent, and the agent can do any of which, then the agent is free. The consequence argument discussed below is particularly useful in objecting to the possibility of this kind of freedom.

Another way to conceptualize freedom requires that the agent be the ultimate source of the action. Accounts in this vein - like Robert Kane's Libertarian view which I discuss later - argue that for an agent to have performed a free action it must originate from the agent as the ultimate source of this action. The agent is free when the agent is responsible for the decision or action rather than some automatic or coerced happening. The idea is that responsibility and agency has to originate in the agent that was not in turn caused by something outside the agent. Incompatibilist

determinists like Pereboom endorse this characterization of freedom maintaining that if there are free agents, they would have to be the ultimate source of their actions. However due to empirical considerations already discussed and the consequence argument to follow, they think that although a possibility, this kind of free will does not obtain in our world.

2.1 The Consequence Argument

The standard argument against free will stands in two parts according to Van Inwagen (1986). The mind argument, is roughly the argument that if determinism is not true then indeterminism is true which means there is randomness at the level of fundamental physics. If this is the case then our actions would be uncaused/random and as such clearly not free. The mind argument however is a threat to libertarian accounts so I explore indeterminacy in considering libertarian free will later on.

The argument from the consequences of accepting determinism on the other hand is known as the consequence argument. The consequence argument has been neatly formalized by Van Inwagen and others in many ways but the following argument is a particularly strong formulation.

1) If Determinism is true, then the conjunction of the laws of nature L and the entire state of the world at an earlier time P_0 entails the entire state of the world at a later time P
2) It is not possible that the agent A can raise his hand at time T and that P be true
3) If (2) is true then if A could have raised his hand at T this would mean A could have rendered P false
4) If (3) and If the conjunction of P_0 and L entails P, then A could have rendered the conjunction of P_0 and L false
5) If A could have rendered the conjunction of P_0 and L false then A could have rendered L false
6) A can't render L false
7) If determinism is true then A could not have raised his hand at T

This argument is logically sound and contains six premises in support of the conclusion. Let's look at the premises.

Premise 1 is simply the definition of determinism. The entire state of the world at an earlier time, plus the laws of nature necessarily produces the entire state of the world at a later time.

Premise 2 follows from the fact that if the agent had raised his hand at T then the entire state of the world (captured by 'P') would have been different than it in fact was at T. The entire state of the world could not thus be accurately described by P if the agent raised his hand.

Premise 3 simply makes explicit that if the agent could have raised his hand at T then he could have made it such that P would be false. If the agent had the power to raise his hand at T then he had the power to render the proposition P (which describes the entire state of the world) false.

Premise 4 relies on what seems to be an analytic principle that if A can render P false, and L entails P, then A can render L false. If L entails P then a denial of P is a denial of L.

Premise 5 relies on another analytic principle. If P_0 is a true proposition about the state of the world which was true before A was born, and A can render the conjunction of P_0 and L false then A can render L false.

Premise 6 needs an argument for the fact that agents can't change the laws of nature to suit their ends. Van Inwagen produces such an argument. He argues that if an agent can render a law of nature false then it is clearly not a law of nature. Ergo, no agent can render a true law of nature false. At times we might think we have settled on a certain law of nature and call it such, but if evidence comes to light that falsifies said law then we were clearly wrong in thinking that we were dealing with a law of nature.

The force of the argument is that if determinism is true and you could have done otherwise, you could have rendered false the laws of nature. None of us have such powers and as such we could not have done otherwise. This is a powerfully intuitive argument against any view of freedom and responsibility that requires genuine alternative possibilities in characterizing responsible agents. By virtue of eliminating alternate possibilities, the consequence argument also rule out ultimate source conceptions of freedom. If an agent is determined by factors beyond his control then he cannot be the ultimate source of his actions as they stem back to factors that preceded his birth.

2.2 Living with Hard Determinism

Although hard determinism doesn't require a positive account of freedom, if its arguments are to be taken seriously, the hard determinist is hard-tasked to persuade people to part with a common-sense understanding of life with free action. A number of vitally important social constructs which do a great deal of work in society all seem to be dependent on the existence of free will. The most difficult pill to swallow for the hard incompatibilist is that if their view is true then nothing anyone has ever done, including Hitler's death camps, has ever really been anyone's fault. There is a strong intuitive pull to believing what seem to be fundamental human experiences and our moral outrage is one such seemingly fundamental human phenomenon. An attractive hard incompatibilist view must find some way to ameliorate these rather unhappy consequences.

Pereboom provides such an account in his aptly named *Living without free will* (2001) which makes the case for the palatability of the way of life in a world where people know hard determinism to be true. In this world we would have no concept of freedom and would no longer think of vengeful anger as justified and the focus would be on the rehabilitation of criminals rather than on their punishment based on the understanding that it wasn't their fault that their causal history has landed them in trouble. A proper discussion of this attractive hard incompatibilist position is outside the scope of this paper. My focus is to defend a more attractive view which also accepts determinism but does not require such drastic departure from our normal way of life.

3 Indeterminism & Libertarianism

Libertarian incompatibilists think that given the unquestionable reality of genuine human freedom, its incompatibility with determinism shows not that we can't have freedom, but rather that determinism must be false or that the causal picture it describes must be re-explained. They prefer to go with our human experience of freedom and trust this intuition in rejecting determinism. This is a highly intuitive position to defend as it fits with our subjective experience of free will and with all of our normal social practices like holding people morally responsible for their actions. There is a serious tension to be explained though, between the information coming from the sciences which describes a mechanistic world and our subjective experience of genuine freedom. It is the Libertarians task to reconcile the fact that our world seems to operate mechanically with the purposive nature of our experience of free will. This motivates libertarians to defend accounts of freedom which involve humans having something special about us that allows us to influence an

otherwise determined system, as this would explain both the mechanical nature of reality and our experience of freedom. Libertarians are 'prime movers unmoved' which allows that we have genuine free choice to do as we please, unaffected by the otherwise deterministic system.

3.1 Indeterminism

Libertarians can't reject determinism outright on the basis of our subjective experience of free will because the alternative would be indeterminism. If the world was entirely indeterministic it would be causally random. In a random/indeterminate world it seems impossible to flesh out any account of how we could be free, as actions would be uncaused and would not seem to belong to the person performing them. In such a world it would be the case that agents would have no control over their own behaviour and agents would be constantly surprised by their own actions. The mere existence of many ways in which a situation could unfold is not enough to allow for *genuine* alternate possibilities. Genuine alternative possible courses of action need to be such that they could have been in some way related to the agents' beliefs and desires.

Some libertarians argue that the development (and huge theoretical success) of quantum mechanics gives us reason to believe that the world isn't entirely causally determined. Our previous best science, Newtonian physics, described the interaction of fundamental particles at the micro level in the same way we describe phenomena at the macro level which is visible to the naked eye. The laws of causation could be depicted by the analogy of balls on a pool table and the predictable deterministic interactions that occur between them. For example, if a ball moving at certain angle and velocity collides with another ball causing the other ball to move away at a certain angle and velocity, then in every case where those precise conditions occurred, the result would be exactly the same. If the fundamental particles (whatever they are) behave in this way, it strongly suggests that determinism is true. In Van Inwagen's terms, if we observed the entire state of play at T_1 and at T_2 we would see that T_1 in accordance with the laws of nature, in a Newtonian Physics world, entails T_2.

However with the advent of quantum mechanics, the fundamental particles are now believed to behave probabilistically rather than deterministically, which some take to be a falsification of determinism. They might argue (by analogy again for simplicity's sake) that if we have the entire state of the world at T_1 and the entire state at T_2, we could not say, given T_1 and the laws of nature in a quantum world, that this entails T_2. Given indeterminacy of the fundamental

laws of nature, it seems rather that T_1 and the laws of nature would instead entail a variety of states of the world at T_2. This encouraged libertarian theorists to take advantage of what has been called the 'quantum gap' in positing theories of freedom in which humans can get involved in causation by influencing causation at the micro level of each quantum event.

For the purposes of this discussion of determinism with regard to responsible agents it doesn't matter whether determinism or indeterminism is true. The first reason to support this claim is that the pool table analogy fails to capture that according to quantum mechanics the average of all interactions at the micro level actually maps the outcome predicted by Newtonian physics at the macro level. So at the micro level there might be a lot of variability, but at the macro level once this variability has averaged out, matter acts in a manner that tends very closely towards Newtonian laws[1]. We do not get pool balls shooting off at strange angles due to quantum indeterminacy. At the level of neurobiology which is the seat of our consciousness (assuming the mental supervenes on the physical) the smallest working parts, as far as we know, are brain cells. Although small to the naked eye, neural cells are constituted by so large a number of fundamental particles that the way matter interacts at this scale would average out to map Newtonian predictions anyway[2]. This is a speculative point but worth pointing out nonetheless.

The main reason that an appeal to quantum mechanics doesn't automatically give the libertarian alternate possibilities and so freedom, is because so long as everything including the mental is *determined* by the events of the past, it doesn't matter whether the fundamental particles interact deterministically or indeterministically. If we could observe the entire state of the world at T_1 and applied the indeterminate laws of nature, we could grant for the sake of argument that we might see some variability at T_2. However it can still be said that any of the various states of the world at T_2 are nevertheless *determined* by the state of the world at T_1 combined with the quantum laws of nature. Whatever state happens to be produced at T_2 by indeterminate causation,

---

[1] The fact that matter acts as per Newton's laws at the macro level of objects big enough to manipulate in all practical purposes is why we teach Newtonian physics in schools.

[2] A counter point in this regard is to argue that according to chaos theory it is possible that some small quantum event in the mind is capable of changing the course of the entire system. Some philosophers think that this is the 'gap' in the closed causal picture through which the human mind can exert some control over the brain in the physical world (Kane, 2001).

it still seems (without any influence from outside the physical) that we are determined by mechanistic causation. So for the purposes of this debate, determinism is the idea that the mental is determined, in whatever way, by the physical and that the physical is governed by regular laws of nature.

In order to take advantage of the quantum gap to spell out an account of libertarian freedom there needs to be a convincing argument of *how* the agent can apply influence in the quantum gap. So Libertarians must provide some theory of how an agent might control indeterminacy in order to exercise freedom. Agent causation is the idea that human beings can get involved in or influence the otherwise deterministic causal process in some way, and has a considerable literature surrounding it. A problem with this kind of agency is that it seems this always boils down to some ethereal self which decides between options, which effectively excludes beliefs, desires and the agent's causal history. This makes decisions seem rather detached from the agent and renders decisions just as random as with uncaused choice.

3.2 Kane's Libertarianism

Robert Kane (1998, p. 5), a prominent modern day libertarian advances a more nuanced libertarian position. He defines free will as "the power of agents to be the ultimate creators and sustainers of their own ends or purposes". He introduces the concept of Ultimate Responsibility and argues that free will is better captured as agents being ultimately responsible for their actions rather than an exclusive focus on whether agents could have done otherwise. In this view agents are under normal circumstances determined to do as they do in line with their beliefs, desires and causal history. So under everyday circumstances they have no access to alternate possibilities in their actions. What makes them morally responsible for these determined acts is that they have been responsible for forming their character at crucial moments in their causal histories. If agents can be held responsible for forming the character which determines their everyday actions, then the agent can be seen to be ultimately responsible for these actions.

At these important moments of character formation agents are torn between competing courses of actions which have moral implications. It is at these moments that an appeal to quantum indeterminacy allows for Kane to postulate that this decision can be made by the agent. The brain has been shown to have the ability to parallel process information, so this allows it to genuinely consider and prepare to choose two different courses at the same time. Each course of action could

have been sufficiently caused by the agent's beliefs desires and causal history had there been no competing course of action. The decision at these moments results in 'self-forming actions' which determine an agent's character and become part of his causal history going forward.

The indecision and mental struggle which occurs at these times of character formation are caused by competing viable courses of action which are pitted against one another. Chaos theory is invoked to explain the experience of struggle when encountering difficulty choosing among competing options which are genuine possibilities for the agent. The idea is that we are able to influence the probabilistic interactions of small quantum events and this sets off a kind of knock on effect resulting in our experience of mental struggle. At times like these we are forced to make character forming decisions because we know that the path we choose will inform our character. When decisions are of moral significance, agents with good character choose morally and can be praised and those who choose immorally can be punished fairly.

Kane's libertarianism explains a great deal well. Agents are the ultimate sources of the decisions, which make them proper targets of morally reactive attitudes like blame and praise. They have genuine alternative possibilities at the crucial time, which makes them responsible for their character making them free in the eyes of those who require that free agents could have done otherwise. Kane's libertarian freedom fits nicely into an otherwise deterministically caused world making use of the quantum gap in a brilliantly intuitive way. Whilst relying on indeterminacy at the relevant point, all decisions are explained in terms of the agent's beliefs and desires which resolves the complaint that indeterminately caused actions are detached from agents' beliefs and desires. Kane doesn't need to worry about arguments like the consequence argument because he is not restricted to determinism and can get around worries about the fixity of the past with his use of indeterminism.

However, at the crucial moment of character formation when an agent must decide between two courses of actions to produce a 'self-forming action', Kane's view seems to run into trouble. At this point something must make the difference or choice between courses of action. Either of the two possible courses of action must be selected by the agent. Kane talks about the effort involved at these times on the part of the agent, but there are many ways to object to this.

If the effort involved is the agent choosing between two possible courses, then we are back to the problem of agent causation as the agent must get involved in the causal chain of events in some way that he is uncaused but is able to cause. Here we worry again that uncaused choice is

arbitrary and doesn't really have much to do with the agent's beliefs, desires causal history or genetic propensities.

If the agent isn't causing the selection, then one of the two courses of events will simply happen. A 'decision' is made in the selection which isn't really one that was caused by the agent. In this case the crucial self-forming actions which make the agent responsible for the formation of his character are determined by blind chance, which makes it seem again like the agent isn't really in control or ultimately responsible for his actions.

If the effort Kane's agents are applying to make the selection at these moments in the crucible of character formation makes the difference, and this effort isn't the kind of ethereal influence traditional libertarians evoke, then I argue that Kane's view reduces to an entirely determined view. If the efforts made on the part of the agent in these moments are determined by all the salient factors that pertain to the agent like his beliefs, desires causal history or genetic propensities, then those too are determined and their effects on the selection between the two courses of action are determined.

So it seems that to avoid agent's character being formed at these crucial times according to chance adoption of one or the other viable courses of action, it seems then that Kane must posit some sort of agent causation – or event causation as he prefers – to make the difference. It is difficult to accept that this kind of agency is plausible given what we know from our best science. Must we accept some sort of substance dualist self which decides between viable options for the agent? I think that this is where a crucial dilemma occurs that hinges on which theory of mind is assumed. Although a polarizing issue in this regard for those concerned with the mind debate, this is not a topic for my present focus.

Casting human free will as the only exception to an otherwise remarkably consistent scientific worldview seems dubious and results in mysterious accounts of causation. Science doesn't prescribe the laws of nature, it just describes what it observes. If the libertarian were correct that humans have the ability to break into the causal chain with some sort of causal influence, this would be the only breach of the laws of nature we know of. It is extremely difficult to accept that human freedom is the only instance of such mystical powers. Discussion of indeterminacy and Kane's libertarian position has served to fill out the libertarian incompatibilist part of the landscape of the free will debate and provides at least some reason to prefer a compatibilist account which I present later on. I consider Robert Kane's sophisticated account as

a viable option for those who think otherwise, but following Fischer and Ravizza think that the mental does supervene on the physical in some way and so I cannot accept Kane's libertarianism.

4 Compatibilism

The consequence argument strongly suggest that agents could never have done other than they did in deterministic worlds. I have argued that there are good reasons to believe that our world is deterministic, or that this is at least a plausible possibility. There is a serious tension here in that free will seems obviously to exist in our world despite hard determinist claims. If we were to accept that free will doesn't exist, this would require large scale social change. There needs to be very good reason to prefer such an account over a view which explains the same phenomena (a deterministic world) but requires less departure from our normal constructs. In contrast to incompatibilist views, compatibilism seems therefore to be a more attractive position in that it allows us to accept a deterministic scientific worldview, the truth of which would not require us to radically restructure society given the lack of free will as an operating assumption. Classical compatibilists and their modern stalwarts have been concerned with developing accounts which make freedom to do otherwise (or some amended version of such freedom) compatible with determinism. There have been countless attempts to save some form of the ability to have done otherwise (McKenna, 2009).

Early compatibilist attempts to get around versions of the consequence argument focused on finding ways to interpret what we mean by free. The most natural first suggestion was that agents are free when they are physically unimpeded from exercising their will (Hobbes, in McKenna 2009). The focus here is on the action rather than the will itself. For example, we are free insofar as we are free of chains or physical coercion. However, counterexamples to this type of early compatibilist position are fairly obvious. Any act performed under hypnosis, psychological illness or mental manipulation will do to dispense with this kind of account because agents are completely physically unimpeded yet they are still not free.

Some such as Hume (1748) might argue for a conditional analysis of freedom. Conditional analysis of free will plays on a slightly altered idea of 'could have done otherwise'. The consequence argument seemed to persuade these philosophers that they lacked the power to have actually done other than they did in fact do. As such the argument was that when we say that an agent could have done otherwise and was therefore free, what we mean is not that they have

magical power over the past or the ability to change the laws of nature, but rather that they had the abilities in that particular situation to have done otherwise given slightly different causal inputs from the past. Conditional analysis views are happy to accept that in lieu of determinism the agent could not have done otherwise given that exact set of conditions and causal history, but that the agent would have done otherwise given a slightly different past, and that this makes her free. Conditional analysis accounts distinguish between those actions that the agent could have performed if she had so wanted, with those she could not have. So long as she was free to do what she could have wanted then she is free.

Conditional analysis ran into trouble because there are cases where agent's ability to choose to act in certain ways are constrained. For example, psychological abnormalities which prevent agents from choosing in certain ways can provide counterexamples to condition analysis:

An agent is asked to choose between two puppies, one with a black coat and one with a blonde coat and it seems completely open that she can choose either. She chooses the black puppy and thinks she could have chosen either, but simply preferred the black one. However unbeknownst to her she could not have wanted otherwise due to a latent phobia of blonde puppies.

It would be a lifetime's work to adequately describe classical compatibilism. I have shown that the consequence argument is a powerful argument against the ability to have done otherwise and it seems that attempts to save any kind of forking paths conception of freedom is bound to fail if the consequence argument rules out genuine alternative possibilities. I think it more profitable to abandon this project in favour of a new compatibilist direction stemming from the main lesson from the consequence argument, which is that determinism rules out access to freedom of the sort which requires genuine alternatives.

5 Routes to Semicompatibilism

There is a distinct shift to be made in our thinking from here on out as Semicompatibilists do not think freedom to do otherwise is the relevant condition for freedom or moral responsibility. Instead, the concern given the possibility of determinism is how agents can be meaningfully morally responsible and therefore free in this sense, quite apart from whether they are free to have done otherwise. Freedom or agency according to semicompatibilist accounts is simply the kind of agency that is required for moral responsibility and not the kind classical compatibilists defend.

There is more than one way to argue for semicompatibilism, my main focus is to provide a defense of Frankfurt's famous cases which, if successful, show that it is possible that agents can be morally responsible even if they could not have done otherwise. If a defense of moral responsibility without the freedom to have done otherwise can be made, this will open the door to an account of morally responsible agents who could not have done otherwise – in other words, responsible agency which is consistent with determinism. But first, I will briefly consider PF Strawson's pioneering work as well as Daniel Dennett's character example in support of semicompatibilism as it would not do semicompatibilism justice to give the impression that this kind of account hinges on Frankfurt cases alone.

5.1 Strawson's Route to Semicompatibilism

Strawson (1962) in *Freedom and Resentment* sees the claimed incompatibility of determinism and freedom in an interestingly different way. He thinks that libertarian freedom is best characterized by the phrase 'obscure and panicky metaphysics' and would prefer not to endorse such a view. Considering views which accept the truth of determinism (determinists and compatibilists) he refers to the former as 'pessimists' and to the latter as 'optimists' about freedom and thinks that we should prefer optimistic accounts.

As his starting point Strawson points out two things:

1) We do not know the truth value of determinism
2) The facts as we know them point to the existence of freedom and morality

However the fact that social practices rely on responsible agency alone does not seem to be sufficient evidence to ground their existence. We cannot make the argument from what is to what must be. The pessimist, in Strawson's terms, is entitled to ask how freedom justifies moral responsibility. The pessimist then argues that the kind of freedom we need to justify moral responsibility is the kind that is not compatible with determinism, that is, the freedom to have done otherwise. And so we have the problem of the apparent incompatibility of freedom and determinism. Strawson thinks that we should work from the facts as we know them rather than speculate on the presently unknown truth value of determinism.

Strawson emphasizes the importance and intricately interwoven nature of morally reactive attitudes like offence, gratitude, resentment, love, forgiveness, affection and praise in the fabric of society. He asserts, and it seems obvious that he is right, that these features of our human

experience are commonplace. He talks about the large vocabulary that exists in order to explore and relate these features of our experience to one another. The idea is that as human beings we are fundamentally moral creatures that automatically see other humans as targets of moral reactions. As we grow up we are given a moral education which makes us see ourselves as targets of moral attitudes of others and to see them as targets in return.

When we encounter a morally reactive attitude, say resentment, there are mitigating factors like 'he had to do it' or 'she had a gun to her head' which are the kind of considerations that result in a relaxation of the relevant morally reactive attitudes. In these cases we are often persuaded that in the actual case the agent was not in possession of the right sort of conditions under which we would want to hold them morally responsible. In fact holding people responsible in certain circumstances is inappropriate no matter what the deed. Strawson makes it clear that in these cases we are persuaded to allow that the particular deed or harm caused was not the agent's fault, even though at all times the agent remains the apt target for morally reactive attitudes.

A second kind of case where we think mitigating factors preclude moral responsibility is captured by phrases such as 'he wasn't himself' or 'he's a hopeless kleptomaniac' or 'he doesn't know who he is'. In these situations we are persuaded to disregard the agent as an apt target for morally reactive attitudes due to manipulation, coercion, psychological abnormality, or when dealing which small children. We adopt what Strawson calls 'the objective attitude'. We treat people as if they are objectively exempt from responsibility because we understand that they are not free in the right sort of way. An objective attitude takes cognizance of these factors and absolves the agent as an apt target for morally reactive attitudes.

In consideration of the upbringing of a child we can see the difference between morally reactive attitudes and the objective attitude. As the child gets older and more competent at being a moral agent, we move away from treating the child as psychologically unable to ground morally reactive attitudes. We begin by regarding the child only with the objective attitude as nothing a small child does is their fault. Children who have not yet learned how to be appropriately morally reactive are treated with the objective attitude as if they are determined to do what they do. "Oh, they're just children" commonly excuses those in a manner that understands that they are not yet capable of being the unfaltering targets of morally reactive attitudes. As they grow we sometimes treat their behaviour with the objective attitude and at other times treat their behaviour as the apt

target of morally reactive attitudes as we teach them how to relate to other people. It is interesting to see the process comes to fruition when we no longer need to adopt the objective attitude. At this point it is almost as if the child no longer is determined by factors beyond its control, because at this point the child should have learnt how human interaction works.

The upbringing of our children suggests that morally reactive attitudes are compatible with either the truth or falsehood of determinism. In this we see how even if we were to adopt widespread acceptance of determinism, human behaviour would come to model our current interaction with children. People might act out initially if everyone came to believe that determinism is true and we were only allowed to treat people with the objective attitude. In the same way we teach our children to be moral we would all begin by treating one another objectively, but we would see people adopting morally reactive attitudes nonetheless due to the built in reaction we seem to have to moral situations. In this way we could not help but revert back to morally reactive attitudes. This is speculative evidence, but there seems to be a strong likelihood that we are psychologically bound to treat each other as the apt targets of morally reactive attitudes whether or not determinism is true. This suggests the irrelevance of determinism in morally reactive attitudes.

If determinism is true, the question is whether we need adopt the objective attitude towards everyone for everything they do and for everything that has been done? Although this clearly isn't an impossible state of affairs, it most definitely seems not be a practically applicable one due to the aforementioned depth at which morally reactive attitudes are embedded in our human culture. In consideration of a child's upbringing we can see that the sustained objectivity of interpersonal interactions seems psychologically impossible in our world.

In sum Strawson makes explicit the intuition that moral responsibility, and as such some form of free will with which to ground these morally reactive attitudes, exist whether or not determinism is true. He argues that humans are fundamentally responsible for our actions in the reality of our social lives and as such we must satisfy some account of freedom. Human beings are psychologically bound to see each other as targets of morally reactive attitudes which leads us to behave as responsible agents in our social reality no matter what might be happening at the microphysical level of particle physics or the metaphysical level of logic and reason.

5.2 Dennett's Character Examples

Dennett (1984) in *Elbow Room* popularized an account of responsible agency that depicts people as highly complex mechanisms which are capable not only of intentional stances but also of what he calls the 'personal stance', which applies to those who are part of the moral community and so are morally responsible. This account does not require freedom to have done otherwise, as Dennett is quite convinced that human beings are at bottom just highly complex mechanisms which are situated in nature and operate according to its deterministic laws in the same way as all other creatures and objects. Dennett brings to relevance the Humean idea and indeed summons many of the same intuitions as Strawson in that we can only be held responsible for those deeds which are determined by our character and motives, quite apart from whether or not causal determinism is true.

Character examples aim show that we do hold people responsible for at least some instances of actions committed where no alternate possibilities were available, purely on the basis of their character. For example, when Martin Luther broke from the Roman Catholic Church he was well known for saying 'here I stand. I can do no other'. What he means is not that he lacks freedom or responsibility but rather that his character determines that he had no other option but to break from the Vatican. The salient point according to Dennett is that we don't think that Luther should be exempt from praise or blame for this action just because he couldn't have done otherwise. In cases like this neither free will nor morality depend on the existence of alternate possibilities. This is an instance where it is given that alternate possibilities are not available yet we still intuit that Luther was responsible for and free in this action.

From this case we might draw the conclusion that moral responsibility and freedom need not always depend on the ability to have done otherwise. In other words, at least sometimes we act freely and in ways that could be reprehensible or praiseworthy even when we don't have alternate possibilities. However it might be argued that this is only the case because we can assume that at some point Luther made choices where there were genuine alternative possibilities, which made him into the kind of person who could not have done other than break from the Church of Rome. If this is so then 'could have done otherwise' is still a prerequisite to freedom in an indirect sense. We need a stronger version of this kind of case which eliminates all alternate possibilities. This is exactly the project Frankfurt undertakes.

5.3 Frankfurt's Route to Semicompatibilism

Although Strawson and Dennett provide alternative routes to a similar semicompatibilist conclusion, I believe the strongest route is through a defense of Frankfurt (1969) who changed the landscape of the compatibilism debate with his famous cases which show that sometimes moral responsibility and the ability to do otherwise come apart, thus rejecting the incompatibilist intuition that 'could have done otherwise' (referred to by Frankfurt as the principle of alternate possibilities, which is shortened to PAP) is a necessary condition for moral responsibility and free will in this sense. He provides cases in which agents could not have done otherwise but where we still think that they are morally responsible for their actions. These cases usually involve a third party who is able to force the agent to do some act even if they decide against it and thereby eliminates all alternate possibilities. For example:

> Jones is deciding whether to shoot Smith. If Jones chooses to shoot Smith, Black's Frankfurt controller device (say a sophisticated chip in Jones' head which can monitor and control his behaviour directly) will do nothing. However if Jones is about to refrain from shooting Smith, the controller will intervene and make Jones follow through. Jones shoots Smith on his own accord.

So Jones was always going to shoot Smith and could not have done otherwise, yet it still seems that he did so on his own and that he is morally responsible. This suggests that PAP and moral responsibility sometimes come apart, leaving the door open to those who want to push an account of moral responsible agents without the freedom to have done otherwise.

5.3.1 Dilemma Defense

Robert Kane (1985) argues in what has become known as the dilemma defense of PAP, that Frankfurt examples are ineffective in their attempts to show that there are cases where morally responsible agents had no genuine alternative possibilities. Consider a case where an agent is about to make a choice for which he can be judged as morally responsible.

> Say Jones is deciding whether to shoot Smith once again. If Jones chooses to shoot Smith, Black's Frankfurt controller device will do nothing. However if Jones is about to refrain from shooting Smith, the controller will intervene and make Jones follow through.

This is a standard Frankfurt case in which the agent seems to have no alternate possibilities available as Jones is shooting Smith either way. But importantly, if Jones shoots Smith, then

according to PAP he is not morally responsible for this morally dubious act because he had no alternative options available to him. This normally counts as a counterexample as we intuit that Jones is responsible because he shoots Smith in the first case on his own accord without the controller having to force him. But the dilemma defense argues that the Frankfurt controller (Black's device) cannot work to produce this conclusion. Black with his control device faces a dilemma: either determinism is true or indeterminism is true, and on either horn the argument is that Frankfurt cases fail to provide a counterexample to PAP.

Let's start with indeterminism (say for example the kind Kane argues for where agents influence quantum indeterminacy at the appropriate time) and assume that until the moment of decision it is undetermined whether Jones will shoot Smith or not. On this horn of the Dilemma there are two possibilities for Black. On the one hand, if in order to ascertain whether Black should intervene he must wait for Jones to decide whether he will shoot Smith, then whatever Jones chooses (say he chooses to shoot Smith) he had genuine alternative possibilities available at the moment of his decision, as Black's controller has yet to intervene. Once the decision is made Black might still be able to force the shooting behavior, but it is too late for Black to get involved with regard to Jones' choice. Jones has made a free choice (with alternative possibilities) for which he is responsible. So it is clear that on this possibility we do not have a counterexample to PAP. On the other hand if Black uses his controller to preempt the choice by getting involved before Jones has made his choice, Jones seems morally exempt in shooting Smith because Black, through the controller, has made the immoral deed occur, not Jones.

Simply put the indeterministic horn of the dilemma runs as follows. If the controller doesn't preempt the agent's choice, then the agent has genuine alternatives at the moment of choice. So the agent is morally responsible whatever he decides by virtue of alternate possibilities. However if the controller gets involved to preempt the choice, then the agent doesn't actually choose at all because the controller decides, so the agent is not responsible. So it seems on the indeterministic horn that Frankfurt cases do not work to produce the conclusion that moral responsibility and PAP come apart.

On the deterministic horn of the dilemma we assume determinism obtains in the example. In the case where Jones shoots Smith and Black does not intervene, Frankfurt would say that this is a case of moral responsibility without alternative possibilities and so thinks he has a counterexample to PAP. The Dilemma defense argues that this is a question begging assertion

because the issue of Jones' moral responsibility is the very issue at hand. Another way of putting this according Goetz (in Fisher 2011) is that in the deterministic case Black and his device "drop out" and have nothing to do with Jones in the actual situation. So if causal determinism obtains, Black has nothing to do with Jones' decision. It seems then that Jones has no alternate possibilities by virtue of determinism rather than because Black has ruled them out. So to argue that by virtue of Black's lack of involvement Jones is morally responsible without alternate options here is simply to assert this conclusion.

5.3.2 Breaking the Defense

Any account that holds people responsible but does not require PAP must try to explain why the dilemma defense fails or they must provide some other reason to think that moral responsibility does not require PAP. Fischer advances an account which relies fairly heavily on Frankfurt's denial of PAP and is obliged to attempt the former. In response to the dilemma defense Fisher (2011) has to admit that Frankfurt cases do not prove that moral responsibility does not require PAP, but he thinks that the basic insight, or 'moral of the story' as he puts it, can be saved - that 'if causal determinism rules out moral responsibility, it is not in virtue of eliminating alternative possibilities' (ibid, 36).

Central to providing Frankfurt cases that do not fail in the face of the Dilemma defense is the idea that there must be alternative possibilities available to Jones and that these need to be robust *before* Black is required to get involved. The mere existence of any sort of alternate possibilities is not sufficient to count as genuine alternate possibilities as they must be of the right sort. They need to be the sort of possibilities that could ground the agent being morally responsible if they were to take these alternative courses of action.

Fisher doesn't attempt a full defense of the indeterministic horn – perhaps because his own view is built to deal with determinism – but he does offer some thoughts. Fisher's suggestion is that we might set up a case where in order to have genuine alternate possibilities the agent must satisfy some condition which is necessary for alternate possibilities which would be causally sufficient to ground moral responsibility. The case is set up so that if Jones were to have a robust alternative possibility of deciding not to shoot Smith, there would have to be a certain thought which occurs before this decision is made. We stipulate that if a robust alternate possibility were to be taken, then Jones would have to show some prior sign. This gives Black the chance to get involved at the time of this thought, prior to the decision but not prior to the beginning of the

necessary unfolding of that decision given the thought. This requires a sort of no-man's land between some prior sign and the actual decision in which Black can get involved to rule out robust possibilities.

Pereboom (2001), although not a semicompatibilist, provides such a case. I paraphrase his example:

> Joe is deciding whether or not to evade tax. He knows this is illegal but in this particular case he knows the chances of getting caught are slim and he could easily get out of trouble by pleading ignorance. He has the disposition to take such immoral self-interested risks but does not always do so. He is a libertarian free agent which means he can take advantage of indeterminacy in the world to produce actions which are not wholly determined by the laws of nature. He is psychologically bound such that the only way he could fail to evade tax is for moral reasons of a certain persuasive force. He could not decide against tax evasion for any other reason – say on a whim. His libertarian free will allows him to evade tax even if a sufficiently forceful moral reason is available to him.

Our notorious neurosurgeon Black inserts a chip into Joe's brain which is programed to force him to evade tax if it senses a reason of sufficient force forming. In actual fact no such reason forms and Joe evades tax.

This is a case where it is undetermined up until the moment of choice whether Joe will evade tax or not, but importantly, if some reason does come to light with sufficient force to make him consider not evading tax, Black's device will prevent him from not avoiding tax. In this case Black can relax (in his disposition to get Joe to avoid tax) because he knows that the only way Joe is going to fail to evade tax is if some sufficiently forceful moral reason occurs and in this case his controller is programmed to counter this.

Unfortunately I cannot see how this example works, because even in the case where the prior sign (the moral reason of sufficient force) occurs, it is still undetermined after this prior sign whether Joe will in fact choose to evade or not evade tax. It seems even in this case that the preemption occurs too early for Joe to have had robust alternatives at the actual time of decision which Black rules out. Luckily for the semicompatibilist this problem falls to anyone pushing an indeterminate account of responsible agency. I will set this aside now to focus on the deterministic horn of the dilemma.

Fisher provides a more thorough, and I think more successful, account of how Frankfurt cases might successfully counter PAP on the determinism horn of the dilemma. Essentially the complaint was that if causal determinism obtains then Frankfurt controllers become irrelevant to the actual sequence of events and so no inferences about moral responsibility can be drawn from their lack of involvement. A new kind of Frankfurt case must be provided to deal with this complaint.

> Jones is deciding whether to shoot Smith and is standing by Black with his brain control chip at the ready. We assume that determinism is true but we do not assume that this rules out genuine alternative possibilities. So we are agnostic about the existence of alternate possibilities under determinism. In this case we will also assume that Jones will *only* have a furrowed brow if he is going to shoot Smith. So if he is going to refrain from shooting Smith then he will lack this brow movement. Black now knows when Jones is going to shoot Smith and when he isn't. Black can now relax (in his morbid obsession with having Jones kill Smith) because he knows that if Jones has no furrowed brow his control chip in Jones' brain will pick this up and force him to shoot Smith. Jones ends up furrowing his brow and shooting Smith on his own accord.

In this case we have not assumed that there are no alternatives for Jones, for example he could have exhibited some other brow movement involuntarily. But this possibility is, as Fisher (2011) put it, only a 'flicker of freedom' which does not seem robust enough to ground moral responsibility. This is because it seems that such an involuntary movement is not the kind of alternative upon which we can ground moral considerations due to its involuntary nature. So although there are alternative possibilities (which seems to spoil the case in that the agent would be morally responsible whilst satisfying PAP) it seems that these possibilities are only flickers of freedom which are not the robust kind required for moral judgment. The fact that Black has ruled out Jones not shooting Smith combined with the assumption of causal determinism going into the case together rule out all robust alternate possibilities. We have a case where Jones, in shooting Smith, is morally responsible and without alternate possibilities, and so we have at least one case where PAP is not true. The argument from Fisher (2011) can be generalized as follows.

1) Causal determinism obtains and the Frankfurt case of Jones shooting Smith unfolds as above

2) Causal determinism is not assumed to rule out all alternative possibilities. Black and his controller are also not able to rule out all possibilities alone.
3) Causal determinism plus Black, his device and his murderous dispositions rule out all alternative possibilities for Jones
4) If Jones is not morally responsible for shooting Smith, then it is not by virtue of the *mere* fact that he was not free to choose otherwise
5) If causal determinism rules out Jones' moral responsibility for his choice to shoot Smith it is not in virtue of its eliminating alternative possibilities – if in fact it does eliminate alternate possibilities

Why should premise 3 require that Black and causal determinism rule out all alternative possibilities? Black, having seen a furrowed brow, will not do anything because he knows causal determinism and the furrowed brow will result in the shooting of Smith.[3] However if Black were not involved in the example, then an alternative possibility is present and Jones could have done otherwise. It is only by virtue of Black on standby combined with causal determinism that all alternate possibilities are ruled out. We may thus still affirm the 'moral of the story' from Frankfurt cases and claim that causal determinism doesn't rule out moral responsibility simply by virtue of eliminating all alternate possibilities.

5.4 Frankfurt's Hierarchical Semicompatibilism

It would be disingenuous to rely so heavily on Frankfurt examples and not to consider his positive thesis at least to show that we might do better elsewhere. Frankfurt (1971) takes an agent to be free in the sense required for moral responsibility when his first order desires result in action which his second order desires endorse.

First order desires are best likened to those you would expect animals to have as they are simple urges to do things. As human beings we have the capacity to 'want to want' which are second order desires about first order desires. According to Frankfurt, an agent is free when he has a second order desire about a first order desire and that it is acted upon. When all three of

[3] We might worry that the furrowed brow and determinism rules out moral responsibility by virtue of the fact that given these conditions Jones is causally determined to do as he does. This however is begging the question, as moral responsibility is at issue here.

these criteria are met the agent has a volition and he can be said to have a free will. This locates the source of the action in the agent, as opposed to some externally caused action, and so the agent can be said to have acted of their own free will.

> For example Jones is still deciding whether to kill Smith. He wants to kill Smith so has a first order desire, he wants to want to kill Smith so has a second order desire, and he kills Smith. In killing Smith based on a first order desire which aligns with a second order desire, he has the volition to kill Smith and so is morally responsible for killing Smith.

Following from his rejection of PAP, Frankfurt's hierarchies are consistent with being entirely causally determined. Although killing Smith was causally determined, Jones did so freely by virtue of his volition to kill Smith and can be held morally responsible.

However, hierarchical views are not without fault and face two main challenges which have limited their success. The first is that it is not clear why second order desires cannot conflict as is a well-known tendency with first order desires. We could have a first order desire to want tea and a first order desire to want coffee and we must choose one as preeminent. The same can be said about second order desires. We can want to want tea and want to want coffee at the same time. If second order desires can conflict then the agent would require a third order desire with which to identify the will. Conflict can occur at the third level and the fourth and as such we have a very real possibility of an infinite regress.

The second problem with hierarchical views is that preeminent second order desires do not necessarily go with free will. In certain cases second order desires can be manipulated so that the agent thinks and feels they are acting freely in accordance with their volition, when in actuality they are being manipulated to feel that way. An example case of this is the willing drug addict.

> Harry is addicted to heroin. He has the first order desire to take the drug. He endorses this first order desire with a second order volition in taking the drug. Harry is also convinced that his second order desire to want to want the drug is his own desire and is not caused by the drug.

The problem for Frankfurt in cases like these is that there is no way to separate out whether Harry is free in wanting to want the drug or if he is caused by the addictive properties of Heroin to want to want the drug. It seems clear that Harry is not free to have a second order volition of his own accord due to the addictive and compelling nature of this destructive habit. Although a fairly brief dismissal of Frankfurt's positive thesis in light of the many attempts to save this view, his

contribution in the rejection of PAP will serve to support an arguably less problematic view in the section to come.

6 Reasons Responsive Semicompatibilism

I have discussed three routes to semicompatibilism with a distinct focus on Frankfurt examples as they show that if causal determinism does rule out moral responsibility then it is not by virtue of eliminating all alternative possibilities. This marks a distinct shift in the way compatibilists can think about free will. If determinism doesn't rule out freedom by virtue of the lack of alternate possibilities, then we are at liberty to propose new accounts of how we might be responsible agents without PAP. Although Frankfurt's hierarchical account advancing such a view has seen some difficulties, I think reasons responsiveness compatibilism developed by Fischer and Ravizza (1982, 1998 & 2011) is a worthy successor. They argue that an agent is morally responsible and so free (although not to do other than she actually did) when she acts with a responsiveness to the right sorts of reasons, under the right sorts of conditions in the actual sequence of events that leads to action.

6.1 Guidance Control

After establishing in the previous section that we do not require genuine alternative options to be responsible agents, I will now provide Fisher and Ravizza's crucial distinction between guidance and regulative control, following which, I will look at Fischer and Ravizza's positive thesis of responsible agency without alternate possibilities.

Compatibilism has been concerned with providing an analysis or interpretation of alternate possibilities which is consistent with determinism because *prima facie* it seems that in order to be free it must be the case that the agent could have done otherwise. Fisher (1982) based primarily on Frankfurt's rejection of PAP, makes a useful distinction between the two types of control over action that an agent might have which allows semicompatibilists a new angle on the debate. Regulative control is the kind of control an agent would have if she had the ability to choose freely between alternate possibilities and really could have chosen either, as with the forking paths conception of freedom. It can also be seen as the power or ability to have done otherwise or the avoidability of what was in fact done.

However this project has run into serious opposition from incompatibilists. The consequence argument has shown that this kind of control is not compatible with determinism. I have also argued that appeals to indeterminism, as with Kane's libertarian view, are unlikely to help in this regard because so long as our actions are determined by the microphysical interactions, it does not matter if causation is determinate or indeterminate. There are compatibilist accounts which try to advance regulative control accounts of freedom but it seems that this is a hopeless task. I have made the case that attempt to explain how an agent can have either forking path or ultimate source control is bound to appeal to some sort of mysterious involvement on the part of the agent, which I take to be a libertarian endeavor.

I have argued in line with Fisher that Frankfurt's rejection of PAP has been successful. This allows Fisher to posit another kind of control, which he calls guidance control, which does not require regulative control over action (PAP). Guidance control is entirely consistent with an agent being morally responsible for an action even though she couldn't have done otherwise. Guidance control however is a slightly weaker construct than regulative control in that it only requires that the agent is part of the causal process that results in the free action. Consider these two cases of control which are analogues of the cases provided by Fisher (2011), to make clear the difference between guidance and regulative control.

> Fisher is driving his car. Everything is as you would expect from this respected philosopher; his car is working well, he is driving normally and wishes to make a right turn. As a result he indicates and carefully executes a right turn, guiding the car to the right. We assume that if he wanted to go left he could have indicated left and gone left. He has control *over* his vehicle. Insofar as he is guiding the car we shall say that he has 'guidance control' over the car. We make no special assumption about determinism or malicious neurosurgeons and we assume that he also has 'regulative control' in that he has the power to have made a left turn when in fact he made a right turn.
>
> Consider a second case:
>
> Fisher guides his car to the right and everything is working properly when going to the right. Unbeknownst to Fisher the steering apparatus isn't working normally for any other maneuver than the right turn he is currently undertaking. If he tried to turn left instead the car would veer off and execute the identical right turn he actually is making. Since Fisher is turning right

the steering apparatus appears to work normally and the car goes exactly where he is guiding it, so his guidance control of the car is the same as in the first case.

In this second case Fisher is guiding the car in some sense to the right. He doesn't cause it to veer off in this direction (perhaps due to a seizure which, with a normal steering apparatus, would normally make him crash or veer out of control), he is simply guiding the car to the right and as such has guidance control. He has control of the car in some sense, but he lacks regulative control over the car. This case is much like the Frankfurt cases discussed earlier, but here it serves to tease apart the two notions of control.

This case should (like the Frankfurt cases) summon the intuition that we do not need to possess regulative control to have moral responsibility, or the appropriate kind of freedom. The fact that the car isn't under his regulative control does not exempt Fisher from fault if some ill consequence were to issue from the car going right, as the lack of regulative control here has no part to play in his actual practical reasoning at the time. We would worry that this case isn't as tightly sealed off from alternate possibilities, but that work has already been done with the discussion of the Frankfurt cases. This case is simply to make explicit the difference between guidance and regulative control.

Accounts of free will that utilize guidance control admit that agents might be entirely determined to do what they do, which makes it difficult to imagine how they could possibly be responsible agents. In placating this concern, semicompatibilists pick out the mechanism of free action which actually plays out within agents before each free action and specify that for the an agent to have committed an action for which he is morally responsible it must satisfy the conditions of whatever mechanism is picked out. This type of compatibilist has to find some way of defining a set of criteria which if met, must render it necessary that the action performed by the agent is morally responsible and so free. Accounts like this are best described by Pereboom (2001) as "causal integrationist" accounts'. I will now explain how Fisher and Ravizza's account works.

6.2 Conditions for Freedom

Reasons responsive compatibilism stems back to the Aristotelian idea of man as a rational animal. Humans are animals in that we are part of the natural world but what sets us apart from the rest of the species that we know of is that we react to reason. We do things because they advance our ends and our behaviour is usually for some reason or many reasons. We think of

reasons as causally important and as such we judge people based on the reasons upon which they act. The kinds of reasons we react to in moral situations are the kinds of reasons which are minimally moral, which explains why we do not include young children and smart animals as proper members of the moral community, allowing them more leeway. We also assume that the reasons involved in moral judgments are the kind that are intelligible to a hypothetical third party, which explains why we do not judge the criminally insane by the same standards as the rest of society. There are three requirements to reasons responsive compatibilism.

6.2.1 Being moderately reasons responsive

The general requirement is that agents must be responsive to reasons and they must act according to some set of these. According to reasons responsive semicompatibilism, agents are 'receptive' to reasons, which means they are receptive to the range of reasons available to the agent which pertain to any prospective act. Agents are 'reactive' to reasons which is the idea that agents can act according to reasons which are sufficient to move them to action. To be exposed to the widest possible range of reasons, agents are readily receptive to reasons. Whilst agents might be receptive to any number or reasons they need not react to all the reasons that are available to them. Only some sets of reasons are strong enough to move us to action.

This distinction allows Fisher to separate out a strong sense of reasons responsiveness, which is too strong and clearly goes beyond what is required for moral responsibility. An agent is strongly reasons responsive when an agent's deliberation and reasoning about an action results in action, and if there were sufficient reasons to have acted differently, the agent would have been receptive to these reasons and would have acted differently by the same mechanism that actually caused her to act as she did. By 'sufficient' reason for alternate actions, Fisher means reasons which would justify alternate action. This condition is too strong because agents who did not, or could not have become aware of a reason sufficiently strong to have done otherwise, or if such reasons were apparent and the agent was able to disregard them, then we could not hold such agents morally responsible. For example

> John is deciding whether he is going to pay a fine or avoid paying. He has strong reasons to pay the fine as he has a court summons in his hand and would most certainly like to avoid going to jail. However, if his mother were sick and needed money for an operation, he would have found these reasons sufficient to have not paid the fine.

John might not have had this alternate reason, say he was an incredibly selfish person and could never have summoned reasons to avoid having to pay that fine. If reasons responsiveness requires a reason sufficient to have done otherwise, then we couldn't hold John responsible.

Moderate reasons responsiveness is less stringent as it only requires that agents consider reasons for and against an action and their reaction to *those* reasons allows us to pass moral judgments about their action in moral situations. The 'moderate' requirement restricts attention to the reasons an agent is actually receptive to when making the decision.

Fisher's aim is to provide an account that maps our normal intuitions about when agents are morally responsible and free. Consider the case of Harry the drug addict who has the irresistible urge to take Heroin. We do not hold him morally responsible for taking the drug and do not think he is free because we are sensitive to the fact that Harry is physically compelled by his brain chemistry to take the drug and as such he is not reasons responsive. His action issues from some physical mechanism which does not pass through his reasons responsive mechanism and as such he is not a responsible agent.

6.2.2 Taking Responsibility

Agents must take responsibility for their reasons responsive mechanism in that they should recognize when their reasons and motivations are the causal source of action. Even if an agent is entirely causally determined to have done as he did, he can be said to have acted freely if he endorses the causal mechanism and reasons which caused the behaviour. Say for example a man decides to steal a beer (let's assume he's not an alcoholic). He has reasons for this which, along with his desire for the beer motivate him to steal said beer. It doesn't matter whether there are alternate possibilities, like not stealing the beer or asking a friend to get him one instead. What does matter is his reactivity to the range of reasons to which he is receptive which results in the theft. By virtue of his having stolen the beer in a way which was reasons responsive and the mechanism which caused this theft being his own, the agent acted in a manner which was of his own will and for which he is morally responsible.

Agents must see themselves as targets of the moral expectations of others as can be gauged by their morally reactive attitudes. This advances the Strawsonian intuition that humans are fundamentally moral creatures in their interactions. In seeing ourselves and others as

appropriate targets of morally reactive attitudes we ground our moral responsibility in our interactions.

Mechanism ownership is subject to a number of provisos which are part of the epistemic conditions which must be met in order for the action to count as free. These include agents being aware of what they are doing and what the consequences of their action will be. This requirement precludes all sorts of freedom-undermining conditions like cases in which the agent's mechanism is addled by any third party manipulation like hypnosis, brainwashing or direct brain control.

6.2.3 Checking Based on the Evidence

Checking for these conditions should be based on the evidence as is available to those involved in each actual scenario. This is admittedly a subjective aspect of the reasons responsiveness. Although we do not like to admit subjective aspects in precise philosophical theorizing, this aspect of reasons responsive compatibilism maps nicely onto our experience as epistemically limited beings. We cannot always know everything involved in every scenario. For an extremely skeptical example we could claim that we do not know for certain that there isn't an evil demon/god who makes us think we are reacting to our own reasons when in fact he is covertly manipulating us to think this way. We are compelled then to revert to the best evidence available when making our judgments.

6.3 Criteria for responsible agency in sum

In order for agents to count as responsible according to reasons responsive semicompatibilism, they must satisfy three conditions.

1) They must be responsive to reasons and reactive to some set of reasons in performing some action.
2) They must be the apt target of morally reactive attitudes and this requires the agent to have taken responsibility for the reasons responsive mechanism which produces the action.
3) The agent must meet the first two criteria based on all the evidence available to them.

6.4 Advantages over Frankfurt's Hierarchies

Reasons responsiveness semicompatibilism doesn't require second order desires for the endorsement of the actual mechanism of reasoning which results in action. In taking responsibility a reasons responsive agent need not appeal to the specific endorsement of an effective first order

desire by a second order desire. Rather a reasons responsive agent's mechanism is theirs by virtue of their awareness and acknowledgement that it is theirs. This allows Fisher and Ravizza to avoid the trap of infinite regress which Frankfurt encountered.

Reasons responsiveness compatibilism also deals with addiction cases in a simpler more intuitive way. Consider again the case of Harry the heroin addict:

> Harry is addicted to heroin. He has the first order desire to take the drug. He endorses this first order desire with a second order volition in taking the drug. Harry is also convinced that his second order desire to want to want the drug is his own desire and is not caused by the drug.
>
> In the Frankfurt case there was no way to separate out whether it is the chemically addictive nature of the drug which causes Harry to want to want the heroin. Reasons responsiveness semicompatibilism would not have this problem as Harry cannot be the apt target of morally reactive attitudes. He lacks capacity to take ownership of the mechanism which causes him to take the drug because the mechanism is a physical mechanism which is outside of his control due to heroin's chemically compelling properties.

7 Objections & Responses

7.1 How might we claim to have done otherwise?

Reasons responsiveness semicompatibilism admits that responsible agents could not have done otherwise. It seems that there is a bit of explaining to do with regards to our normal way of understanding the phrase 'could have done otherwise'. What might we mean when we utter this phrase if reasons responsiveness semicompatibilism is true?

Consider the actual causal chain leading up to an action to see how this can be understood. An agent decides to take a right turn at a traffic light. She is moderately responsive to reasons which are her own and there are no special circumstances which need precluding. If we were to say that the agent did turn right but could have done otherwise, what we mean is that she did turn right but if she were to have other reasons (perhaps she saw a traffic jam to the right and could better go left) then she could have acted otherwise. We restrict attention to the reasons responsive mechanism which produced that action. If other reasons had been in play then the agent may well have done otherwise. Understanding 'could have done otherwise' like this maps onto what we

normally mean by the phrase. We do not normally consider determinism or its preclusion of alternate possibilities when we normally utter this phrase.

7.2 The Consequence Argument

I have set out the consequence argument and shown that it is a powerful objection to classical compatibilism or any view which advocates what amounts to a garden of forking paths view on freedom. After having argued for reasons responsive semicompatibilism I should show how this view deals with what is possibly the strongest rendition of the standard argument against free will.

The consequence argument argues that if determinism is true, the past is fixed, and we lack the power to change the past, then we could not have done other than we have done and will do. Semicompatibilist accounts are immune to any such arguments which aim to undermine freedom to have done otherwise simply because semicompatibilism doesn't rely on freedom to have done otherwise. It has been shown through discussion of Strawson, Dennett and Frankfurt that the principle of alternate possibilities is not a necessary requirement for moral responsibility and the subsequent discussion of how responsible agency can be captured by reasons responsiveness semicompatibilism has shown that the freedom required to be morally responsible in this way does not rely on alternate possibilities. In fact proponents of reasons responsiveness compatibilism are at liberty to accept the consequence argument, but they need not.

7.3 Phil the newly converted hard incompatibilist

Mele (2000) tells a story about a good and gregarious previously-compatibilist philosopher called Phil who gets converted to hard determinism through far too much involvement with philosophers who are convinced this is the correct way to understand the problem of free will. In line with his newfound dedication to hard determinism he no longer believes that he is an apt target for morally reactive attitudes and so believes that he cannot be held morally responsible. By stipulation in this case Phil is incorrect in his belief in hard determinism and the world is actually one in which there are agents, moral responsibility, and people are fair targets of morally reactive attitudes.

Even though Phil now denies free will and responsibility he nevertheless carries on with his life as usual. He continues to donate to charity, raise his children to be morally sound people, and

behaves in all ways as a moral person would. He does this not because he thinks that he is metaphysically justified in doing so, but does so because living in a world where morally reactive attitudes abound produces favorable social outcomes. Pretending to be moral is convenient for Phil.

Mele suggests that agents need not see themselves as the apt targets of morally reactive attitudes and that it is entirely possible that agents can in all respects behave morally without ever satisfying the 'taking responsibility' condition required by reasons responsive semicompatibilism. The point is that it seems we can have agents who are morally responsible yet do not see themselves as having taken responsibility. It is thus conceptually possible that moral responsibility and taking responsibility come apart. So the question is whether taking responsibility is a necessary condition for moral responsibility?

To drive home the point that morally responsible reasons responsive agents need not require morally reactive attitudes, Mele considers a community of emotionless people (who lack the capacity to be apt targets of reactive attitudes) that nonetheless satisfy all the other conditions of the reasons responsive account and are moral in their behaviour. They meet all criteria for the account to count them as moral when they clearly are not members of our moral community. If this is the case then it seems reasons responsiveness does not capture all the relevant features of freedom/moral responsibility.

Fischer and Ravizza are happy to concede that Mele has pointed to the most difficult part of the theory to explain due to its subjective nature. However their response (2000) deals well with this complaint. It seems that Mele has suggested the impossible. It does not seem to be possible for Phil to believe that all things considered (including metaphysical considerations from his change of heart in switching to hard determinism) that reactive attitudes are unjustified, whilst still having such attitudes. It does not seem possible to both genuinely have morally reactive attitudes and at the same time believe that they are all things considered unjustified. Consider the morally reactive attitude of resentment. It doesn't seem possible to genuinely resent someone whilst at the same time believing that they do not deserve said resentment and that you cannot justifiably hold such resentment given your metaphysical considerations.

If Mele is taken to be saying that Phil somehow sees himself as a target for morally reactive attitudes by others yet he doesn't endorse his own mechanism for acting the way he does, it seems that Phil is somehow disconnected from his behavior. Phil is like the sailor of his ship, but he is

passively going along with the motions because he knows his rudder is broken. He does what he must do but doesn't really feel that he is freely doing what is being done. In this case he isn't really taking responsibility. He is only going through the motions and so is not free according to the account, so we do not have a counterexample here.

The worry about a community of robots or emotionless beings that nevertheless operate morally does not really worry Fisher and Ravizza because they feel that they have not specified that the Strawsonian focus on morally reactive attitudes is the only conception of morality that could work with the reasons responsive mechanism. They endorse this particular view because they find that it fits well with the evidence available from this world. But in a possible world in which such emotionless beings exist, we might adopt a different scheme of moral responsibility, like a moral ledger which only focuses on deeds of right and wrong.

7.4 Problems with reactivity

Mele (200) thinks that Fisher and Ravizza's reasons responsive mechanism too readily accrues blame where in fact there should be no blame. He conjures an interesting counterexample in which an agoraphobic man called Fred who has not left his house in ten years, and due to this crippling fear has not attended his daughter's wedding in the church next door. In this case we would not think him morally responsible or at fault. However if we look at the reasons responsive mechanism and more specifically the reactivity clause, Mele contends we would have to hold him responsible because (given the possibility of fire) he could have had a reason strong enough to leave the house and attend.

Agents are regularly receptive but weakly reactive to reasons; this means that they can easily have and consider many or all of the reasons available to them, but they are only weakly reactive in that not all reasons for doing things are acted upon. Mele's contention is that it is actually pretty easy to imagine some reason that would have moved Fred to leave the house. This is easy enough. In a possible world with the same laws as Fred's we have a similar case as described above except in this world there is a raging fire in Fred's house on his daughter's wedding day. It turns out Fred is more afraid of fire than of leaving the house and has to overcome his agoraphobia through great effort. He does so and stumbles out the house into his daughter's wedding. This serves as a reason strong enough for him to have left the house and as such we can hold him morally responsible for not having left his house on his daughter's wedding day. If he

could've left for the fire then it can be said that he could've left for the wedding, and as such he is morally responsible for not having left for the wedding. If a theory blames people where it shouldn't, then there has to be something wrong with the theory.

Fisher are Ravizza (2000) solve this problem simply by pointing out that just because Fred is an apt target for morally reactive attitudes, this does not entail that he is in fact morally responsible. They point out that agents can be moderately reasons responsive and therefore responsible for morally neutral acts (like whether to turn left or right at a traffic light). The morality or immorality of Fred's failure to leave the house is determined based on the actual reasons he had and this is left open to those in the situation to decide. Moral debate turns on the reasons Fred had available. Considering the situation holistically, including the agoraphobia, the morally reactive attitudes of the people in this situation would probably not find Fred morally responsible.

Fischer and Ravizza have opened the door for a theory of praiseworthiness and blameworthiness but have not advanced one. Although this seems to be a detracting factor I take this to be a point in favor of the reasons responsiveness compatibilism. Morality might accurately be described as a universal human trait, but it is well-known fact that particular accounts of morality differ from culture to culture. It makes sense to state a universally applicable account of how we might ground moral responsibility and freedom metaphysically but leave the specific reasons and cultural consideration open for debate about each actual situation.

7.5 Kane

For Kane (1999) a problem with reasons responsiveness compatibilism is that it seems very much like humans lack the kind of access to our subconscious minds to have any chance of meaningfully checking based on the evidence as to whether we are being manipulated. It seems that there is no way for the agent to tell if they are being fiddled with by their subconscious minds. Thus we cannot meaningfully claim to be source of our actions.

Additionally if we were to look at the agent's causal history it seems that there is very little difference between the types of coercion that are not permitted by this account and the kind of coercion that simply must be at play given the accounts' possible acceptance of determinism. An agent's causal history is operating in controlling him beneath the level of conscious recognition required for taking responsibility in any meaningful way.

Fisher and Ravizza might concede that there is a great deal of evidence to show that our unconscious minds do in fact manipulate us at times but that this is part and parcel of accepting determinism. Kane is insisting that the agent must be source of her action in some way. Semicompatibilism doesn't require that the agent must be the ultimate source in taking responsibility. In a determined world everything is determined, so the concept of ultimate responsibility does not make sense with regard to semicompatibilism. We cannot be the ultimate source of our actions if we are entirely determined beings. Reasons responsive semicompatibilism doesn't claim special circumstances where the agents can get involved in making choices in their causal history. It is saying that agents who are morally reactive creatures that behave responsively to reasons when taking responsibility for their actions can be seen as agents and judged according to the moral standards of whatever moral situation they are in, quite apart from whether they can be said to be ultimately responsible.

On the issue of checking for manipulation based on the evidence, I now consider the strongest possible objection in this regard.

7.6 The 'Four Cases Argument' for incompatibilism

Pereboom (2001) makes an argument known as the four case argument for incompatibilism which aims to show that if we are determined by the past then we cannot be morally responsible. Pereboom asks the reader to consider four different cases where an agent, Mr. Plum, kills Mrs. White. Each case is alike in that an agent is manipulated into behavior but with each case the amount of manipulation is decreased. Pereboom argues that covert manipulation in the first three cases generalizes to an entirely normal causally determined fourth case.

In the first most obviously unfree case an agent, Mr. Plum, was created by neuroscientists who can control his brain directly but he has no idea that this is the case. These neuroscientists can control everything about Mr. Plum including the practical reasoning that leads to his murdering Mrs. White. We would not want to hold Mr. Plum morally responsible for directly forced behaviour.

In the second case Mr. Plum is no longer under direct control of the neuroscientists, but when he was made it was programed that his practical reasoning would at some time in the future be such that he would undertake to murder Mrs. White and he performs the deed without knowledge that the behaviour was programmed. This too seems obviously unfree and holding the agent responsible seems unfair.

In the third case Mr. Plum is a normal human being except that at a very early age he was rigorously indoctrinated such that by virtue of this conditioning he will at some stage kill Mrs. White. He does not have any memory of this indoctrination and kills Mrs. White based on his own practical reasoning and completely endorses this action as his own. This case is less obvious, but still seems unfree and we wouldn't hold him morally responsible as it was not his fault he had been indoctrinated. If the compatibilist wishes to say that he is responsible, then some relevant difference must be pointed out that separates the second case from this one. If he is not responsible in the second case then it seems he shouldn't be responsible in the third.

In the fourth case Mr. Plum is a normal human being who lives in a physicalist, determinist world and he kills Mrs. White out of his very own practical reasoning. He is unaffected by any strange manipulation except that he lives in a deterministic world and he is caused by his past to behave as he does.

Pereboom argues that there is no relevant difference between the first three cases and the last, and that we should therefore not hold Mr. Plum responsible for this murder. We cannot hold Mr. Plum responsible and that seems to rule out moral responsibility where determinism obtains. It seems that if determinism is true it entails that we are all effectively manipulated into acting as we do. Pereboom worries that some might claim that the relevant difference is that in the first three cases the manipulation was achieved through some outside agent. To further drive the point home that there is no relevant difference Pereboom asks the reader to imagine that the first cases of manipulation were carried out by some insentient, non-purposive machine that spontaneously happened to cause Mr. Plum to have Mrs. White expunged. Additionally the first three cases are effectively counterexamples to the reasons responsive view. These are powerful objections to which Fisher and Ravizza must respond.

Fisher (2011) admit that simply being moderately reasons responsive is not enough to ground moral responsibility as Mr. Plum satisfies this condition in the first three cases where he clearly isn't morally responsible. Reasons responsiveness compatibilism requires that the mechanism which produces the action must belong to the agent, in that Mr. Plum needed to have taken responsibility for his reasons responsive mechanism at some point. A reminder that reasons responsiveness semicompatibilist responsible agency requires these three conditions:

1. Mr. Plum must see himself as an agent in that his choices must affect the world and he must believe that if he had different reasons he'd have acted differently. He must be reasons responsive.
2. Mr. Plum must see himself as a fair target of morally reactive attitudes which precludes freedom undermining conditions.
3. The aforementioned conditions have to be based on the evidence.

These requirements do not seem to help Fisher and Ravizza against the four cases argument. Mr. Plum meets the first requirement in that he is reasons responsive and he has definitely seen a change in the world due to his actions – that is, the death of poor Mrs. White.

The second requirement seems to be where Fisher and Ravizza have some hope of a response. Manipulated agents cannot take responsibility for the mechanism which produces action because the reasons are not their own. If agents haven't taken responsibility for their mechanism of action then they cannot be held responsible. From outside the situation as third party observers we can see that Mr. Plum isn't a fair target of morally reactive attitudes, but according to the information Mr. Plum has access to, there seems no reason for him not to see himself as a fair target of morally reactive attitudes. To him this seems to be a perfectly normal case and he would (assuming he is psychologically normal) agree that he was guilty of the crime and morally responsible.

The third requirement is also met. According to the evidence available Mr. Plum can draw the conclusion that he meets requirements 1 and 2. We should note at this point that Mr. Plum isn't required to know all the details about how his mechanism produced the death of Mrs. White. In cases of ordinary practical reasoning we don't know the details that go on at the subconscious level but we take responsibility nonetheless. The evidence available to Mr. Plum in all these cases is that he is a normal person acting under normal circumstances and that there are no factors precluding his moral responsibility.

Fisher and Ravizza's account does not seem to be able to handle the four cases argument. It seems that they need to explain why the four cases argument cannot make the claim that there isn't a relevant difference between the first three cases and the last in order to save moral responsibility from the threat of determinism. There does not seem to be any transfer of non-responsibility principle at play, so finding fault in the premises of the argument seems a lost cause. But I think there is a way to cast some doubt on the four cases argument. First I deal with the

direct attack on the possibility of moral responsibility in a deterministic world and then I attempt to provide some reasons why we should not consider the four cases argument a threat to reasons responsive semicompatibilism.

First it must be noted that the four cases argument cannot assume that determinism rules out moral responsibility. Seeing as moral responsibility is at issue, to assume this would be question begging. So to be explicit, we are now assuming determinism is true and that it is up for debate whether or not the lack of a relevant difference between the first three cases rules out the compatibility of determinism and moral responsibility. It seems like the assertion that there is no difference between the first three and the last case is less safe than Pereboom might think.

Consider the four cases from a godlike third party's perspective. With all the relevant information we can see that Mr. Plum is not morally responsible in the first three cases due to manipulation. In the fourth case there is no manipulation. He is caused by determinism to do exactly as he does but this does not automatically amount to manipulation. As mentioned above, Pereboom suggests that we could strengthen the argument if we remove agency from the story. Instead of neuroscientists being responsible for Mr. Plum's creation, programming and indoctrination, these acts of manipulation are carried out by insentient computers completely spontaneously. This makes the manipulating effect of determinism seem even more analogous to the first three cases of manipulation. However, there is still a difference in that there is manipulation in the first three cases and none in the last.

So one way of arguing against the four cases argument is to flat-footedly insist that there is a relevant difference between the first three cases and the last simply by virtue of the fact that there is manipulation over and above causal determination. This point is strengthened by the fact that in a deterministic world, which may or may not contain morally responsible agents, the difference between the first three cases and the last seems obvious to any third party observer with all the information about the situation at hand. In the first three cases we intuit that Mr. Plum is definitely not morally responsible because there is clearly manipulation which makes him kill Mrs. White. In the last case we know he is made to kill Mrs. White by determinism, but we don't know (if the four cases argument is not asserting the incompatibility of moral responsibility and determinism) whether this is enough to say that Mr. Plum is manipulated.

If it were assumed that determinism rules out freedom or moral responsibility directly then the fourth case is clearly one of manipulation and we could not hold Mr. Plum responsible. However

since it is not assumed, it is then possible that even though in the fourth case Mr. Plum is entirely caused by determinism to act as he does, he is not manipulated and can still be responsible. So if there is a difference, it is then a question of whether determinism rules out moral responsibility directly. This turns on what we might count as manipulation and whether or not determinism itself counts as manipulation. It is up to semicompatibilist theories of responsible agency to make a case for the compatibility of determinism and moral responsibility. I have shown that reasons responsiveness semicompatibilism is an attractive option in this regard.

Whether or not a difference can be found, reasons responsive semicompatibilism still seems to be in trouble considering the four case argument. In the four cases argument the criteria set out by reasons responsive semicompatibilism will find Mr. Plum morally responsible in the first three cases when we have very strong intuitions that he is not. In each of these cases Mr. Plum is made to kill Mrs. White through a processes of manipulation and we intuit that whatever he has done, it cannot be his fault if he was manipulated. If a theory of free will hopes to be successful it must set out the conditions under which an agent is morally responsible. If these conditions are met it must be necessarily true that the agent is morally responsible and so free. Reasons responsiveness doesn't pick up on the fact that Mr. Plum is not a responsible agent in the first three cases.

I want to argue that reasons responsiveness semicompatibilism need not accept these three cases as counterexamples because these cases assume that moral responsibility must be objectively justified but human kind lacks the power to justify morality in this way. The four cases argument hinges on the fact that information is concealed. In the first three cases it is stipulated that the agent is unaware of factors that would normally preclude moral responsibility if they were to come to light. So there is this asymmetry between what we can see as a third party observer and what the agent making the decisions is aware of. A godlike observer intuits that Mr. Plum is not morally responsible but those ignorant of the manipulation (including Mr. Plum himself) are convinced that he is morally responsible. I think this asymmetry explains why we might be persuaded by the four cases argument when we shouldn't.

As human beings we are limited by our epistemic access to all the factors that might be manipulating us, a fact which Fisher and Ravizza are sensitive to. If reasons responsive compatibilism is true, agents are reasons responsive and take responsibility based on the evidence available to them. We could not have access to all the relevant information about a situation unless we were godlike creatures. Thus we can only take responsibility in a subjective manner due to the

fact that we are epistemically limited beings. I agree with Pereboom that there is no way we would be able to take responsibility for our mechanism of the springs of our actions in an objective manner which would pick up such covert manipulation as in the first three cases.

But this is the problem. The four cases are not cases that are relevant to a world in which agents are epistemically limited. Moral responsibility is thus not the kind of thing which can be decided on objectively by agents who cannot possibly have access to all the relevant information. Any theory of free will which doesn't construe responsible agency as something which requires perfect knowledge of all the facts involved will fail to 'pass the test' presented by the four cases argument. The cases are constructed in such a way that the relevant information is unavailable to the agents and so the agents acting on the evidence available cannot help but fall prey to the completely covert manipulation. Only godlike creatures with full epistemic access could possibly take responsibility based on all the evidence and because there are no godlike humans (I assume) there can be no responsible agency if objective morality is a requirement.

However I still do not think that this means that we should abandon compatibilism and accept that there are no responsible agents. I think that if we cannot show that the four cases argument is itself at fault we would still do better to accept that in some cases reasons responsiveness semicompatibilism might, on account of the subjective element, wrongly accuse an agent based on insufficient evidence about what is going on in the actual sequence of events. In this case Mr. Plum would certainly not argue that he is not guilty for his murdering Mrs. White in court. As far as he is concerned he is guilty in all four cases. In the actual sequence, if Mr. Plum or a third party were to learn of the neuroscientists' covert manipulation then according to the evidence he would not be guilty in the first three cases and (according to reasons responsiveness semicompatibilism) would definitely still be guilty in the fourth case because he would be a reasons responsive agent who is a fair target of morally reactive attitudes. This again maps our normal understanding of responsible agency. We hold people accountable based on the best evidence we have. Unfortunately for Mr. Plum, unless this evidence comes to light, in our world, he would face the brunt of moral responsibility for his murdering Mrs. White in all four cases. A reality of our world many wrongly accused will attest to.

The four cases argument highlights the weak point of reasons responsiveness semicompatibilism in that it contains a subjective element which can be exploited by arguments such as Pereboom's four cases. However if any theory of free will is to capture our normal

understanding of responsibility it cannot require objective proof that agents are not being manipulated because such proof is not accessible to us. I think that the four cases argument is assuming that we cannot be morally responsible if we do not have unlimited epistemic access to all the information involved in a situation. I think that a theory of morally responsible agents like human beings need not have such a strong requirement. Humans are epistemically limited beings and so if any moral theory required godlike knowledge of all the factors relevant to each particular situation, it would not be a theory of any practical use to human beings.

8 Conclusion

I have presented a line of reasoning, assuming that our world is physical and that there can be no causation outside the physical, which defends a reasons responsive semi-compatibilist account of free will and moral responsibility. Various conceptualizations of freedom have been considered along the way.

Libertarian accounts, although offering a robust conception of free will in their insistence that agents must at least have genuine alternative possibilities if they are free, require some mysterious theory of causal involvement on the part of the agent. If the world is physical at bottom, I contend that there is no reason to believe that humankind is specially endowed with the ability to manipulate the physical with the mental. However those who are happy with this sort of non-physical causal picture would find a plausible account in Kane's intuitive use of the quantum gap.

Hard incompatibilists' denial of moral responsibility is a serious problem for their view. We cannot ignore free will as it is an ingrained feature of our subjective reality. Even the committed hard determinists must nevertheless continue living their lives assuming, for the sake of harmonious relations with others and in the general practice of their endeavors, that they have freedom to choose. I find no reason to prefer a determinist account over some form of compatibilism which can explain the same deterministic world but can also retain some notion of freedom, albeit relying on a weaker understanding of free. Nonetheless, Pereboom's account provides some insight into how life might actually be better for the social change involved in accepting that we have no free will.

Compatibilism seems to be the only sensible answer given the lack of empirical evidence to falsify a scientific worldview and our subjective experience of freedom. I have advocated for a particular brand of compatibilism which grew from Strawson's focus on morally reactive attitudes

and Frankfurt's landmark denial of the principle of alternate possibilities. Fischer and Ravizza have advanced a theory of moral responsibility which allows us to accept the incompatibilist claim that alternate possibilities are impossible given determinism, whilst still retaining moral responsibility and the kind of freedom which grounds it. Although this account of freedom seems to yield weaker conception of free will in that it doesn't require freedom to have done otherwise, it is preferable to incompatibilist accounts which either require a breach of our scientific understanding of causation or the complete denial of freedom and responsibility. Finally reasons responsive semicompatibilism is an extremely attractive position in that it accounts for the social reality of freedom and responsibility no matter what the, presently unknown, nature of determinism turns out to be. Our status as morally responsible agents is grounded in our practical reasoning quite apart from whether determinism turns out to be true.

9 Bibliography

Baumeister, R., Mele, A. & Vohs, K., (2010) *Free Will and Consciousness, How Might They Work?* Oxford University Press: Oxford.

Carroll, J., & Markosian, N., (eds.), (2010) *An Introduction to Metaphysics*. Cambridge University Press: Cambridge.

Dennett, D., (1984) Elbow Room: The Varieties of Free Will Worth Wanting. MIT Press: Cambridge.

Dennett, D., (1984) I Could not have Done Otherwise: So What?. *The Journal of Philosophy*. 81(10). p. 553-565.

Conee, E., and Sider, T., (2005) *Riddles of Existence, A guided tour to Metaphysics*. New York: Oxford University Press.

Fischer, J., (1982) Responsibility and Control: An essay on moral responsibility. *The Journal of Philosophy*. 79(1). p. 24-40.

Fischer, J., (2011) *Deep Control: Essays on Free Will and Value*. Oxford: Oxford University Press.

Fischer, J., and Ravizza, M., (1998) *Responsibility and Control: A Theory of Moral Responsibility.* Cambridge University Press: Cambridge.

Fischer, J., and Ravizza, M., (2000) Responsibility and Control: A Theory of Moral Responsibility by John Fischer and Mark Ravizza reviewed in *Philosophy and Phenomenological Research*, 61(2) 467-480.

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

Hume, D., (1748) *An Enquiry Concerning Human Understanding.* Grin: Verlag.

Kane, R., Fischer, J Pereboom, D., & Vargas, M., (2007) *The Grate Debates in Philosophy*. Oxford: Blackwell Publishing.

Kane, R., (2002) *The Oxford Handbook to Free Will*. Oxford University Press: Oxford.

Kane, R., (1999) Responsibility and Control: A Theory of Moral Responsibility by John Fischer and Mark Ravizza reviewed in *The Philosophical Quarterly*, 49(197) 543-545.

Kane, R., (1998) *The Significance of Free Will*. Oxford University Press: New York.

Mele, A, (2000) Responsibility and Control: A Theory of Moral Responsibility by John Fischer and Mark Ravizza reviewed in *Philosophy and Phenomenological Research*, 61(2) 447-452.

McKenna, M., and Russell, P., (eds.), (2001) *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment".* Ashgate: Surrey.

McKenna, M., (Winter 2009 Edition) "Compatibilism", The Stanford Encyclopedia of Philosophy. Zalta, E., (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/compatibilism/>.

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

Strawson, P., (1971) Freedom and Resentment and other essays. Routledge: New York.

Stump, E., (2002) Responsibility and Control: A Theory of Moral Responsibility by John Fischer and Mark Ravizza reviewed in *Philosophy and Phenomenological Research*, 61(2) 459-466.

Van Inwagen, P., (1975) The incompatibility of free will and Determinism. *Philosophical Studies*. 27(3). p. 185-199.

Van Inwagen, P., (1986) *An Essay on Free Will*. Clarendon Press: Oxford.