Remixed Responsibility
Defending a Compatibilist View of Moral Responsibility

By

Tabitha Taylor

Submitted to

Central European University

Department of Philosophy

In Partial Fulfilment of the Requirements for the Degree of Master of Arts

Supervisor: Ferenc Huoranszki

Budapest, Hungary

2016

Table of Contents

Introduction

In this thesis I argue that the best compatibilist theory of moral responsibility for tackling manipulation arguments is one that relies on the concepts of reasons-responsiveness and mechanisms belonging to the agent. In order to show this I will start by presenting a convincing compatibilist motivation (Chapter 1), namely Frankfurt's argument that alternative possibilities are not needed in order to have moral responsibility. I then outline Fischer and Ravizza's compatibilist theory of moral responsibility (Chapter 2) focussing on the basic concepts they use, some of which will be important for my remixed reasons-responsive theory of responsibility, which I outline in (Chapter 5). Some of the concepts, at least in the form that Fischer and Ravizza have them, I will criticise in order to show that they are not needed for an efficient theory of responsibility (Chapters 3 and 4). Having outlined the remixed theory of moral responsibility I am proposing (Chapter 5), I then explore two strong arguments against compatibilist theories of moral responsibility: Derk Pereboom's four case manipulation argument (Chapter 6) and Alfred Mele's zygote argument (Chapter 7). These two arguments have been discussed in great detail by other compatibilists who have proposed different theories of moral responsibility, including Fischer and Ravizza. I tackle these arguments to show that the remixed reasons-responsive theory of moral responsibility is equipped with solutions to these problems and is therefore a strong compatibilist view (Chapter 8).

1. Compatibilism

Whether or not we have free will is something that is considered very important when it comes to analysing peoples' actions and whether or not they are morally responsible for them.[1] Determinism[2] is the idea that everything that happens is entailed by everything that happened before it. The possibility that determinism might be true has led some philosophers to deny that we have free will and thus cannot be held morally responsible for our actions. However, there is strong resistance against this conclusion and many wish to produce compatibilist accounts of moral responsibility, which entertain the possible truth of determinism whilst maintaining that, even if it is true, one can still be held morally responsible for one's actions.

1.1 Free Will and the Principle of Alternative Possibilities

When one is asked to describe what it means to say that one has free will, most often the description includes some appeal to more than one option being open. It is the idea that there is a genuine choice, that there is an alternative possibility to the action actually carried out. When it comes to moral responsibility this idea of being able to do otherwise is essential; can we really hold someone morally responsible for what they have done if they could not have done otherwise? Intuitively, it seems that the answer is "no": we should not hold morally responsible those who really had no other options.

[1] In this paper I will talk mostly of *moral* responsibility because it seems that moral cases are those where responsibility really matters. Nonetheless most of what I say can also apply to cases of responsibility that do not have moral implications or considerations.
[2] By 'determinism' I mean *causal* determinism in its most basic sense, i.e. "*the idea that every event is necessitated by antecedent events and conditions together with the laws of nature*" (Hoefer, 2016).

These considerations become especially important when determinism enters the picture. It seems that if determinism is true, then any ability to do otherwise is erased. Because the truth of Determinism entails that there is only one way for things to go, there is no room at all for alternative possibilities. This puts free will in jeopardy and with it moral responsibility. But do we really need the ability to do otherwise in order to be responsible for our actions? Can free will and responsibility come apart in some way?

Indeed, Frankfurt, in his important paper *Alternate Possibilities and Moral Responsibility* (1969), argues that we do not in fact need the ability to do otherwise in order to be morally responsible. Frankfurt says that the principle of alternative possibilities is false; one may be morally responsible even if one could not have done otherwise (Frankfurt, 1969). This idea has been demonstrated in many different examples of what are known as 'Frankfurt cases'. The general structure of such cases shows that it may be possible that there be such "circumstances that constitute sufficient conditions for a certain action to be performed by someone and that therefore make it impossible for the person to do otherwise, but that do not actually impel the person to act or in any way produce his action" (Frankfurt, 1969).

To see this, consider Frankfurt's own original Frankfurt case. He imagines two agents, Jones and Black. They both want Jones to do a certain action, call it x. Black knows Jones very well and worries that he will waver in doing x. So to ensure that Jones really does x, Black takes necessary steps to make sure that Jones does x if he sees that Jones wavers. These steps might be, say, taking control of Jones' brain and body and thus steering Jones to do x. In fact, Jones does not waver and does x without Black's intervention.

There are two things to notice: first that Jones could not have not done x, i.e. he could not have done otherwise, and second that Jones can rightly be held morally responsible for doing x. In this way, Jones *is* morally responsible *without* having alternative possibilities. If we consider the alternative case where Jones wavers and Black intervenes to ensure Jones does in fact do x, in this case, Jones is *not* rightly held responsible. However, the possibility of this alternative does *not* entail that in the *actual* sequence of events, where Jones does not waver, responsibility is removed for Jones. It is not the case that because Jones would not have been responsible in the alternative scenario, that Jones is not responsible in the actual sequence of events too, as is clear from this Frankfurt case. Frankfurt demonstrates that the relevant phenomena for moral responsibility do not lie in the principle of alternative possibilities (Frankfurt, 1969).

Determinism produces the circumstances that a Frankfurt case considers i.e. where there is only one way that an agent can go. But as Frankfurt demonstrates, this does not mean that the agent is necessarily then not morally responsible. This is because the impetus of the act is still in the agent; their moral responsibility stems from elsewhere, from where exactly shall be described in the following chapters. It is clear that removing the need for the principle of alternative possibilities is a convincing way of confronting determinism head-on. If one can have moral responsibility without needing alternative possibilities, it seems that any theory that incorporates this is shielded from the force of determinism, at least as it is initially conceived.[3]

[3] There are of course other ways in which determinism threatens compatibilist views, even of this strongest kind, but these shall not be discussed here.

2. Fischer and Ravizza's Theory

In their book, (Responsibility and Control, 1998), Fischer and Ravizza, following Frankfurt, deny the need for alternative possibilities, but their approach is different from Frankfurt's own. There are a number of important concepts that they employ which will be explained in each sub-section here. The concept of mechanisms (2.2) will be examined more closely (Chapter 3) and both this concept and that of reasons-responsiveness (2.3) will be important for the remixed theory of moral responsibility I suggest. The concepts employed in the ownership of a mechanism (2.4), on the other hand, will be critiqued (Chapter 4).

2.1 Control

Fischer and Ravizza focus on the idea of control, explaining responsibility in terms of different sorts of control. First they distinguish between guidance control and regulative control. This distinction is based on differentiating the actual sequence of events and the full range of possibilities: regulative control is control in all scenarios regardless of which is actual, whereas guidance control is control in the actual sequence of events only (Fischer & Ravizza, 1998, pp. 28-41).

To describe the difference between these two sorts of control consider Frankfurt's Jones and Black case again. Jones is in control of his actions insofar as he does x, but not insofar as he wavers about doing x. If Jones chooses to do x and does not waver, he is in control of the actual situation. On the other hand, if Jones wavers and Black takes over, then in this scenario, Jones does not have control over his actions. In Fischer and Ravizza's terms, Jones has guidance control, but not regulative control: he can control his actions in the actual sequence, but not in the alternative one. So to have guidance control is to have the power to perform an action freely, whereas regulative control is to have the dual-power to perform freely in all possible

events. Indeed, one can have guidance control without regulative control. As may be clear from the application to Frankfurt's case, guidance control is the necessary type of control for moral responsibility, it being the sort that Jones has. So according to Fischer and Ravizza, and in line with Frankfurt's demonstration, the freedom relevant sort of control for moral responsibility is guidance control, which does not involve alternative possibilities (p. 33). This makes their theory an 'actual sequence account' since it focuses only on the *actual* sequence of events and whether or not responsibility can be attributed *there*. We now see how Fischer and Ravizza shield their theory of moral responsibility from the force that determinism has in removing alternative possibilities. Their theory does not require alternative possibilities for moral responsibility: the impetus of the action is in the agent and characterised as guidance control.

In order to describe further the criteria for moral responsibility, Fischer and Ravizza go on to explain how guidance control may be exhibited. It is not enough that the agent merely has guidance control in order to be considered morally responsible; guidance control is a minimal requirement. Indeed, an animal has guidance control of its actions, but we do not hold animals morally responsible for their actions. So there is more to being held morally responsible than exhibiting this minimal version of guidance control.

2.2 Mechanisms

In light of their actual sequence account, Fischer and Ravizza shift from an "agent-based" approach to a "mechanism-based" approach. This entails focussing not on the agent herself, but on the mechanisms in play when the agent acts. In Frankfurt cases, "the kind of mechanism that actually operates is reasons-responsive, even though the kind of mechanism that would operate – that is, that does operate in the alternative

scenario – is not reasons-responsive" (Fischer & Ravizza, 1998, p. 38). The idea of 'reasons-responsiveness', which will be spelled out in more detail below (2.3), contains counterfactuals: *x is reasons-responsive if x can do otherwise when considering different scenarios*. So reasons-responsiveness requires alternative possibilities. This means that the agent herself cannot be said to be reasons-responsive, since she might not have alternative possibilities, but the mechanism that produces her action can. The plausibility of this move will be further discussed in (Chapter 3).

Thus, Fischer and Ravizza shift alternative possibilities from the agent to the mechanisms that lead to the agent's actions. Returning to Frankfurt's case once more, one might say that the mechanism that operates in Jones' action in the actual sequence is reasons-responsiveness, whereas in the alternative scenario where Black intervenes, the mechanism that operates in Jones' action is *not* reasons-responsive. This is because Jones' reasoning is usurped by Black's intervention in the alternative scenario, and the mechanism that operates is not responsive to reasons. Let us take a closer look at precisely what reasons-responsiveness is.

2.3 Reasons-Responsiveness

Reasons-responsiveness is essential for connecting an agent's reasons for action to the action itself. It is not only that an agent must act in accordance with reasons, but also that she must act *because* of reasons (Fischer & Ravizza, 1998, pp. 63-64). In other words, the reasons must be those that motivate the agent to act.[4] Guidance control

[4] Throughout this essay I am mostly referring to "motivating" or "explanatory" reasons, as they are referred to in the philosophical literature on reasons.

requires that the actual sequence have the right sort of connection between reasons and action.

In order to exhibit this connection, an agent must be both *receptive* to reasons and *reactive* to reasons. Receptivity to reasons is having the capacity to recognise the reasons that there are either to do something or not to do it. Fischer and Ravizza favour a strong receptivity to reasons which requires not only that the agent recognise *a* reason to do something (or refrain from doing it), but that she exhibit a comprehensible pattern of receptivity to reasons, i.e. more than just one single reason. One can see this in such cases as Brown and Plezu (1998, pp. 69-70). In this example, Fischer and Ravizza imagine an agent, Brown, who really likes a non-addictive drug, Plezu, which causes one to waste many hours lounging on the sofa enjoying oneself. Imagine that Brown says he won't take the drug if it costs $1000, i.e. the drug's costing $1000 is a sufficient reason not to take the drug. Because of this, he seems to be receptive to reasons. However we later find that Brown, acting on the same mechanism, *would* take the drug if it cost $2000, $3000, etc., i.e. not recognising these prices as sufficient reasons not to take the drug. Here we would say that Brown, *in virtue of his acting on this mechanism*,[5] is not strongly receptive to reasons because he only has one single reason not to take the drug (its costing $1000), and thus does not exhibit a coherent pattern of receptivity to reasons. If he did, he would not take the drug if it cost $1000 *or more*. Brown must act on a mechanism that is strongly receptive to reasons in order to be held morally responsible for his action. Fischer and Ravizza call this strong receptivity to reasons that demonstrates a coherent pattern of

[5] We will see the significance of this qualification later on since an important part of Fischer and Ravizza's theory is that mechanisms must be held fixed in order to analyse them for responsibility.

reasons, *regular reasons-receptivity*. They state that what we want to know is "if (when acting on the actual mechanism) he recognises how reasons fit together, sees why one reason is stronger than another, and understands how the acceptance of one reason as sufficient implies that a stronger reason must also be sufficient" (1998, p. 71). It is also important that the reasons one recognises be minimally "grounded in reality," i.e. that the agent is not in a delusion. So, regular receptivity to reasons "requires an understandable pattern of reasons-recognition, minimally grounded in reality" (1998, p. 73).

Reactivity to reasons is the essential part for the connection between reasons and actions. However, Fischer and Ravizza only require *weak* reactivity to reasons, despite their requirement for *strong* receptivity to reasons. This is because they have a "fundamental intuition that "reactivity is all of a piece"" (1998, p. 73). What this means is that if an agent's mechanism can react to one reason, it can, and should, react to others like it. So, going back to the Brown and Plezu example, if Brown, in virtue of his mechanism, is regularly receptive to reasons, i.e. he recognises that if $1000 is a sufficient reason not to take Plezu, then any price above $1000 is also sufficient not to take Plezu, then he should be able to act on these other sufficient reasons, e.g. if the drug costs $2000. So, if Brown, acting on this regularly reasons-receptive mechanism, takes the drug even though it costs $2000, we can rightly hold him morally responsible for this action. This is because, given that the mechanism would have been reactive to Plezu's costing $1000 being a sufficient reason not to take it, it could have been reactive to Plezu's costing $2000. Hence the connection Fischer and Ravizza suggest between receptivity and reactivity:

> In the case of receptivity to reasons, the agent (holding fixed the relevant mechanism) must exhibit an understandable pattern of reasons-recognition, in order to render it plausible that his mechanism has the "cognitive power" to recognise the actual incentive to do

> otherwise. In the case of reactivity to reasons the agent must simply display *some* reactivity in order to render it plausible that his mechanism has the "executive power" to react to an actual incentive to do otherwise.
>
> (1998, p. 75)

There is a third criterion for reasons-responsiveness. This is simply that it must include recognition of *moral* reasons, i.e. that the mechanism be issued from a *moral* agent. This then explains why we do not always hold children morally responsible despite their minimal reasons-responsive capacities. To be reasons-responsive, then, is to be regularly receptive and weakly reactive to reasons, including moral reasons.

2.4 Ownership of a Mechanism

Fischer and Ravizza's move from agents to mechanisms makes it very important that the relationship between the agent and the mechanism is secure. The way in which they characterise this relationship, as I now outline, is something that I will explore and criticise later (Chapters 3 and 4).

2.4.1 History

Fischer and Ravizza contend that responsibility is an historical notion. This historicity is a feature that helps to decide whether an action can rightly be attributed to an agent, i.e. whether the agent actually owns the mechanism from which the action came. Not only does responsibility require that the mechanism that led to an action be reasons-responsive, but that the mechanism have a genuine historical relationship to the agent.

First let us examine what use history might have for an account of moral responsibility. Fischer and Ravizza sketch a "tracing account" which tries to refine the relevance of history in responsibility such that historical considerations should only be taken into account when they would point to appropriate information about the

present. This account is used to capture such cases as the drunk driver running over a child. It is plausible that at the time when the drunk decides to get in his car, he is not acting from a reasons-responsive mechanism, but nonetheless we still hold the drunk driver responsible for running over the child. This is because, barring circumstances of forced consumption of alcohol, the drunk allowed himself to get too drunk, and can be reasonably considered to know that this would lead to his drunk driving (1998, p. 50). This case can be generalised:

> ...an agent's act at a time T1 issues from a reasons-responsive sequence, and this act causes his act at T2 to issue from a mechanism that is not reasons-responsive… When one acts from a reasons-responsive mechanism at time T1, and one can reasonably be expected to know that so acting will (or may) lead to acting from an unresponsive mechanism at some later time T2, one can be held responsible for so acting at T2.
>
> (Fischer & Ravizza, 1998, p. 50).

In this way, the present action from a non-reasons-responsive mechanism can be traced back to a choice issued from a reasons-responsive mechanism. Thus, history has great relevance to responsibility since without it, it is difficult to explain why, in such cases as the drunk driver, we hold people morally responsible despite their sometimes seeming to act from a non-reasons-responsive mechanism.

The second important element to the concept of history is the genuineness of the agent's relationship to the mechanism that led to her action. In the case just described, the drunk driver has a genuine historical relationship to the mechanism that issued in his action, despite the mechanism not being reasons-responsive. But what about when the mechanism *is* reasons-responsive and we still don't want to hold the agent morally responsible? Such a scenario might occur in manipulation cases, such as hypnosis, where the agent is not held responsible because she was in some way manipulated into doing what she did. Fischer and Ravizza contend that when the reasons-responsive mechanism from which an agent acts is produced through

hypnosis, or some other artificial means, "the mechanism that issues in the relevant behaviour is not, in an important intuitive sense, the agent's own" (1998, p. 197). In the case of an agent who has been hypnotised, the mechanism from which she acts does not have a genuine relationship to her; it is not her own, but has been artificially implanted. It is in this way that an agent can be said not to have a genuine historical relationship to the mechanism from which the action came, and therefore not be morally responsible.

History is an important element of Fischer and Ravizza's theory since it characterises intuitions we might have about cases in which an agent ought, or ought not, to be held morally responsible, where the condition of reasons-responsiveness does not provide enough information to answer the question of responsibility. In the case of the drunk driver, where the agent acts on a non-reasons-responsive mechanism, but we still want to hold him morally responsible, we do so in virtue of being able to trace back from the action to a reasons-responsive mechanism. In the hypnosis case, on the other hand, the agent acts on a reasons-responsive mechanism, but we *don't* want to hold her morally responsible. This can be explained by the fact that the reasons-responsive mechanism on which the agent acts is not her own because there is not a genuine historical relationship between the mechanism and the agent.

As is clear, the agent's ownership of a mechanism is important for Fischer and Ravizza's theory. To further spell out the notion of the ownership of a mechanism, Fischer and Ravizza propose conditions that an agent must fulfil in order to 'take responsibility'.

2.4.2 Taking Responsibility

'Taking responsibility' is not something like a speech act where you say, "I am responsible for x"; instead it is a set of dispositional beliefs. These include (1) viewing oneself as the source of one's actions, (2) viewing oneself as an apt target for the reactive attitudes,[6] and (3) that these beliefs (i.e. these ways of viewing oneself) are based appropriately on one's evidence (so as to rule out delusions) (1998, pp. 215-230). If these beliefs are in place with regard to a certain mechanism, the agent can be said to take responsibility for the action that the mechanism lead to.

Fischer and Ravizza describe how these beliefs might appropriately come about, i.e. through a normal "moral education". This includes treating children as if they were full moral agents, even when you consider them not to be, in order that they might learn to be one. For example,

> a young boy, overcome by excitement, tears open the presents belonging to the birthday girl, despite having been instructed in the proper etiquette. We might well correct him and show the customary signs of indignation, even though we are well aware that the child is not yet fully responsible.
>
> (1998, p. 208)

By teaching a child to understand that they are the source of their actions and that they are accountable for them, one teaches them to be a moral agent, a fully fledged person. Indeed, Fischer is willing to deny personhood to those who do not develop in the proper way: "… take a baby before the baby becomes a moral agent. Scientists come and start manipulating the baby's brain. I would say that that baby never becomes a person, because the baby never develops into a moral agent" (Fischer, 2000, p. 417). Without getting into a discussion about personhood, it is clear that

[6] This idea about 'reactive attitudes' comes from (Strawson, 1962), which is discussed by Fischer and Ravizza (1998, pp. 5-8).

Fischer takes the moral education of children very seriously. Indeed, it is an essential part of the taking responsibility condition. The account of moral education he relies on is a very intuitive sense of a 'normal' or 'typical' case, which I will also make use of.

We can now summarise Fischer and Ravizza's conditions of responsibility: an agent is responsible for a certain act, if the action comes from a reasons-responsive mechanism that has a genuine historical relationship to the agent and the agent can rightly be said to have taken responsibility. Only together are these three criteria sufficient for responsibility.

3. Another Look at Mechanisms

Though the move from agent to mechanism allows for a strong compatibilist view, one might argue that it is problematic. What does it mean to say that a *mechanism* is reasons-responsive? Due to the counterfactuals contained in the concept of reasons-responsiveness, it might look like reasons-responsiveness cannot apply to a mechanism, but only to an agent. The agent is the one with (or indeed, without) alternative possibilities, not the mechanism: being receptive to a coherent pattern of reasons and being reactive to any and all of the set of reasons that are "all of a piece", contain counterfactuals *for the agent*. In this way it begins to look as if Fischer and Ravizza do not in fact avoid alternative possibilities, but rather hide them in the mechanisms. There are three questions that I wish to address. First, is it plausible that these counterfactuals can be successfully subsumed by mechanisms? Second, how can we individuate mechanisms? And third, if it is the *mechanism* that is reasons-responsive, why then is it that the *agent* is responsible? I will attempt to answer these questions by further teasing out the details and implications of Fischer and Ravizza's theory, as well as adding a new concept, namely that of the general reasons-responsive cognitive power, which will be further explored in (Chapter 5).

3.1 Are Fischer and Ravizza Hiding Alternative Possibilities in Mechanisms?

The first issue is whether a mechanism can do the work an agent does in being reasons-responsive. The question boils down to whether or not a mechanism can be independent of the reality of the alternative possibilities; since the agent has no alternative possibilities, how can a mechanism be meaningfully said to have caused the agent to do otherwise in light of appropriate reasons? Indeed, since the alternative

possibilities don't exist for the agent, how do we make sense of them seeming to exist nonetheless for the mechanism?

I argue that a mechanism can be considered as a particular instance of an agent's more general "cognitive power". This cognitive power is a very general ability to be non-specifically reasons-responsive that is then instantiated in a particular mechanism. This idea is similar to Michael Smith's (2003) account of rational capacities. Smith contends that there are rational capacities that can be abstracted away from outside factors, such as Black's intervention in Frankfurt's Jones and Black case. These "intrinsic features" of the agent's can, hypothetically, be applied in many cases, not just the one in which the agent happens to be, and can be understood as the structures that "underwrite[s] the psychological states in general" (Smith, 2003, p. 25). Though Smith's own view ultimately requires alternative possibilities, this notion of isolating cognitive capacities or powers is something that can apply here.

Take the example of Jones and Black from Frankfurt's case again. Jones is generally a rational person and his general cognitive power to be abstractly reasons-responsive works well. In the particular circumstance of considering whether or not to do x, the mechanism employed is specifically about x, the reasons to do it and not to do it, but is still in the general sense, hypothetical: Jones' mechanism would *hypothetically* cause him to respond to sufficient reasons not to do x *regardless of* the actual possibility of Jones' not doing x (remember that in fact Jones cannot not do x because Black will intervene if Jones wavers). In the actual sequence, however, the mechanism leads Jones to do x. The mechanism, as an instantiation of Jones' general reasons-responsive cognitive power, is responsive to reasons that might apply to

doing x, regardless of whether or not Jones actually has all options open to him, i.e. whether or not he really can either do x or not do x.

In this way, the mechanism need not pay attention to the actual options open to the *agent*, it is merely an instantiation of the general cognitive power of abstract reasons-responsiveness in a specific instance, much like Smith's structures of psychological states. Indeed, the possibility of Jones' wavering suggests that he has the reasons-responsive mechanism despite not having the option of putting it into action if the mechanism tries to lead him to not do x. If he did waver, Black would intervene, artificially placing a mechanism to do x in Jones. In the instance of intervention, whether or not the artificial mechanism that Black implements is reasons-responsive, Jones is not responsible for the action that issues from this mechanism since it is not an instantiation of his general reasons-responsive cognitive power.

So an *agent herself* can be abstractly reasons-responsive, insofar as her mechanism *would* lead her to do otherwise in light of reasons, even if the agent herself cannot be said to be able to do otherwise. The agent, though lacking alternative possibilities, has the general cognitive power of abstract reasons-responsiveness. In the actual sequence of events, it is the *mechanism* that is relevantly reasons-responsive.

It is important to stress, however, that the mechanisms are the focus of Fischer and Ravizza's theory, not this general cognitive power, and indeed my interpretation of it (which will be further explained in Chapter 5). R. Jay Wallace's (1996) theory of responsibility focuses on the general cognitive powers an agent possesses, rather than the specific mechanisms that operate in the action. Wallace argues that "reflective

self-control", the essential cognitive power needed for responsibility, is reasons-responsive in a similar sense to Fischer and Ravizza (Wallace, 1996).

The problem with a focus like Wallace's – just on the general capacities and not on the specific, actual-sequence mechanisms that issue an action – is that one is forced to draw strange conclusions like the fact that a person might be responsible for something they do under hypnosis since they do indeed possess the relevant cognitive powers at that time, whether or not they are in action. Wallace says that the fact that the powers are *not* in action can help to explain why responsibility is not attributed in these sorts of cases. But this then means that he must explain what the difference is between exempting responsibility in cases of hypnosis and exempting responsibility in the case of determinism (Fischer, 1996). Thus, by analysing just the actual-sequence mechanisms that actually operate in an action, we avoid any counterintuitive consequences and problems such as these.

Wallace and others[7], on the other hand, criticise the use of mechanisms saying that it is difficult to say precisely which mechanism is operating in any given action and to individuate between them. Let us now examine this problem of mechanism individuation.

3.2 Mechanism-Individuation

Fischer himself admits that "one has to say something about mechanism-individuation" since much of his theory hangs on "holding a mechanism fixed" (Fischer, Responsibility, History and Manipulation, 2000). Fischer and Ravizza's account of mechanisms does not give details about how one might individuate

[7] For example (McKenna, Review of Responsibility and Control: A Theory of Moral Responsibility, 2001)

different kinds of mechanism. Fischer and Ravizza rely on an intuitive method of distinguishing which mechanism is the relevant one:

> It is simply a presupposition of this theory as presented here that for each act, there is an intuitively natural mechanism that is appropriately selected as the mechanism that issues in action, for the purposes of assessing guidance control and moral responsibility.
>
> (1998, p. 47)

Indeed, it is difficult to give an account of mechanisms in the mind without delving deep into philosophy of mind and neuroscience. It might be suggested, however, that one could, in theory, identify a mechanism in the brain if one knew enough about how the brain works.

Michael Mckenna in his (2001) review of Fischer and Ravizza's book, challenges this intuitive concept of identifying a mechanism in the context of "holding a mechanism fixed". As we shall see he also finds problems for Fischer and Ravizza's view with the suggestion that identifying a mechanism can be done through knowledge about the brain. McKenna claims that the notion of 'sameness' that Fischer and Ravizza seem to appeal to might be challenged. He argues that if it were the case that mechanisms were defined by their micro-neurophysiological-properties, then it would be implausible that one could hold such a mechanism fixed (McKenna, 2001, p. 97). This would be a problem for Fischer and Ravizza's account because it would render their "holding a mechanism fixed" notion impossible.

However, it seems implausible to me that such a thing as a reasons-responsive mechanism, as far as it might be intuitively understood, could be reduced to micro-properties without losing much of the meaning of reasons-responsiveness. Reasons-responsiveness seems to be an irreducible macro-property and thus using micro-properties to identify a reasons-responsive mechanism is futile. Compared with a micro-property identification of mechanisms, Fischer and Ravizza's intuitive

approach now seems much more plausible. Indeed, when such things as reasons are appealed to in the context of philosophy, it is difficult to say how they can be reduced to anything more specific than simply something to count in favour of doing something.[8] There doesn't seem to be a coherent way of describing what a reason is in reduced physical terms, let alone micro-physical terms.

Thus, Fischer and Ravizza's appeal to mechanisms is both plausible in itself and lends itself to a more successful theory of responsibility than a general capacities approach like Wallace's.

3.3 How Does the Agent Relate to the Mechanism?

The third issue I now turn to is why the agent is held responsible if it is the mechanism that is relevantly reasons-responsive. This can easily be resolved when we consider again ownership of mechanisms and taking responsibility for the actions that mechanisms issue. As has been laid out, Fischer and Ravizza's criteria for responsibility include more than just a reasons-responsive mechanism. The mechanism must have a genuine historical relationship to the agent and the agent must take responsibility for the actions that it leads to.

This is a nice feature of Fischer and Ravizza's view since it allows for an agent to have physically carried out an action, but to be able not to take responsibility for it owing either to some violation in the genuineness of their relationship to the mechanism that led to the action, or to their lacking the relevant beliefs to be said rightly to take responsibility for the action that the mechanism led to. It is in this sense that, for Fischer and Ravizza, the agent herself and not merely her mechanism is

[8] This is not to say that there are not different kinds of reasons, the distinctions between them being very important, for example, motivating vs. justificatory reasons and subjective vs. objective reasons (Lenman, 2011).

properly considered to be responsible or not. With the addition of the general reasons-responsive cognitive power, which is itself a proper part of the agent, we can think of responsibility stemming from the specific instantiation of this cognitive power, i.e. the reasons-responsive mechanism. Only as a proper part of the agent, to the extent that a cognitive power is part of an agent, can a mechanism be responsible. So agents are those to which responsibility is attributed, via the mechanisms and their general reasons-responsive cognitive power.

I contend that the move from agents to mechanisms that Fischer and Ravizza make is plausible. If a mechanism is thought of as an instantiation of the more general cognitive power of general reasons-responsiveness, and it is still the agent that is responsible, not merely the mechanism, then we can see the plausibility of this move.

4. Another Look at History and Taking Responsibility

4.1 Avoiding Incompatibilism

In his analysis of Strawson's view of moral responsibility (Watson, 2004), Gary Watson presents a problem for compatibilist views that introduce an historical dimension. Watson describes how historical information about people can shape our reactive attitudes and thus our intuitions about whether or not someone is responsible. He argues that incompatibilists will see this appeal to historical factors as an essential part of responsibility, thus supporting their claim that an agent's actions are an "inevitable product of his formative circumstances" (2004, p. 243). Seeing an agent in this way does not lend itself to the idea that the agent is also responsible. Watson contends that one cannot be responsible for circumstances over which one had no control and so if an agent came to be a certain way through events that were forced upon her, then she is not responsible; "It is this principle that gives the historical dimension of responsibility and of course entails the incompatibility of determinism and responsibility" (2004, p. 243). In this way, adding an historical dimension to responsibility in fact supports the incompatibilist's agenda. Watson sees a compatibilist view that has an historical component as "grist to the incompatibilist's mill" (2004, p. 243). It seems that Fischer and Ravizza's reasonably unbounded notion of history does come up against these accusations of incompatibilism. The remixed theory, on the other hand, can more easily avoid them as we shall see in (5.2).

Another way to see this problem is to notice that opening up the floor for history to be a part of responsibility admits the possibility for factors outside of an agent's control to affect their responsibility. As Robert Kane points out, "when one does start thinking historically about responsibility, one is liable to fall into the

clutches of us incompatibilists" (2000, p. 401). This is due to the fact that there is an inherent tracing principle in the incompatibilist's intuition that stems from the thought that, to be responsible, the agent has to be "the ultimate creator and originator" of their decisions to act (2000, p. 401). Kane argues that alternative possibilities creep in when one starts looking at the history behind an agent's actions because it concerns the choices they made in the past that brought them to where they are today. As we have seen, alternative possibilities conflict with determinism, and so Kane suggests that this appeal to history makes Fischer and Ravizza's theory of responsibility once again vulnerable to the force of determinism. Kane questions the difference between a moral education of the sort Fischer and Ravizza suggest and an indoctrination case where the compatibilist intuition is that the agent seems not to be responsible because of her history. The question is how to determine which agents and which actions are apt candidates to apply the tracing principle to and thus exempt responsibility.

Fischer's reply to Kane on this issue is to argue that there is an intuitive difference between moral education and indoctrination and though it is difficult to give a clear account of precisely what this difference is, that does not mean that it is not there. For Fischer, there does seem to be a difference between "mere causal determinism" and indoctrination (Fischer, Chicken Soup for the Semi-Compatibilist Soul: Replies to Haji and Kane, 2000). Fischer maintains that his and Ravizza's view avoids likening causal determinism to such cases as indoctrination through their account of reasons-responsiveness and ownership of mechanisms. For those unsatisfied with this reply, however, as mentioned earlier, the remixed theory avoids such worries by not relying so heavily on this historical notion.

4.2 Not Taking Responsibility

The notion of 'taking responsibility' that Fischer and Ravizza employ in their theory is one that has incurred some criticism. Ishtiyaque Haji (2000) presents a number of cases where agents who, intuitively, are not responsible still seem to take responsibility, and where agents do not to take responsibility yet seem intuitively to in fact be responsible. One of the examples in which the counterintuitive nature of Fischer and Ravizza's 'taking responsibility' condition is most salient is when we take an agent, Ivan, who is a strong believer in determinism and its negative consequences on moral responsibility. Ivan does not see himself as an apt candidate for the reactive attitudes owing to the many incompatibilist philosophical arguments he has rehearsed. According to Fischer and Ravizza's theory then, it seems that we cannot hold Ivan responsible for his actions because he does not take responsibility for them. Haji believes that this is a "troubling result" (2000, p. 398); why should these particular beliefs that Ivan has about determinism affect his responsibility? The problem is not just that the conclusions one must draw about such agents is somewhat counterintuitive, but also that if these particular philosophical beliefs are irrelevant for the specific actions that Ivan might do, and "there are no other intuitively responsibility-undermining factors (like inappropriate manipulation) that infect the etiological pathway of the actions […], then it is not evident why these beliefs should make the difference that Fischer's account implies they do make" (Haji, 2000, pp. 398-399). Thus it seems as though Fischer and Ravizza's conditions for taking responsibility are perhaps not rightly thought to be relevant to an agent's responsibility.

Indeed, one might be able to do without these conditions altogether as I intend to show with the remixed theory. To briefly see how, one can appeal to the general

reasons-responsive cognitive power to solve problems of whether a mechanism belongs to an agent or not. If the mechanism has an authentic relationship to the agent, via the agent's own general reasons-responsive cognitive power, then the mechanism can rightly be attributed to the agent and therefore the agent is responsible for the action that it issues. This idea will be further explicated in (5.2).

4.3 Another Feature of the Ownership of a Mechanism

In reply to a manipulation argument from Todd R. Long (2004), Fischer (2012) suggests that the mechanism operating in an action need not only be reasons-responsive, but also must "exhibit the appropriate capacity to evaluate the new input" (p. 198n28). Fischer contends that an artificially introduced mechanism, even if it is reasons-responsive, must be accepted, as it were, by the agent in order for the agent to be morally responsible for the action that the mechanism produced. He also admits that

> if an "input" is artificially implanted in such a way as to leave it open to the agent (in a reasonable and fair way) to critically scrutinise and reflect on the new input, then this sort of manipulative indiction of inputs *may well be* compatible with moral responsibility…
>
> (Fischer, 2012, p. 197)

According to Fischer, for the mechanism to be truly an agent's own, it need not only have a genuine relationship to the agent and the agent must take responsibility, but the mechanism must also be reflected upon by the agent. So it seems that neither the historical component nor the taking responsibility component of Fischer and Ravizza's theory are sufficient to block Long's argument. Fischer tries to introduce a new element, this evaluative capacity, but perhaps there is a better way to characterise this solution.

One can interpret the requirement for an evaluative capacity in the context of the general reasons-responsive cognitive power I have introduced above. Since it is

the agent's own general reasons-responsive cognitive power that the agent's mechanisms are an instantiation of, then the artificially implanted mechanism is not relevantly reasons-responsive *to the agent*, despite being reasons-responsive in itself. However, if the agent were to reflect on this mechanism through her general reasons-responsive cognitive power, then she might find that it is acceptable to her. If she then acted on this newly accepted reasons-responsive mechanism, though it was artificially implanted, she would rightly be held morally responsible for the subsequent actions issued via the mechanism. As a simple example, take an agent, Amelia, who is wondering what to get her friend Daisy for her birthday. Imagine that Daisy would really like Amelia to get her a bracelet, so artificially implants a mechanism that should lead Amelia to buy Daisy a bracelet for her birthday. Once this mechanism has been implanted, Amelia reflects on it through her own general reasons-responsive cognitive power, and finds it acceptable. In this way she is responsible for buying Daisy a bracelet for her birthday, despite Daisy implanting this mechanism artificially. Notice though, that Amelia's responsibility has nothing to do with the fact that the mechanism was inputted artificially. What is important is that the mechanism was analysed, and thus in some way can be thought of as an instantiation of Amelia's own general reasons-responsive cognitive power. We might imagine that, having thought a bit harder about it, Amelia would have come up with a bracelet as an idea to get for Daisy for her birthday. Daisy's implanted mechanism just sped up this process. Mechanisms must be reflectively monitored by the agent's own general reasons-responsive cognitive capacity in order for the agent to be responsible.

5. The Remixed Theory

Fischer and Ravizza's comprehensive account of reasons-responsiveness explains how best to interpret the notion of guidance control and what it means to be responsive to reasons. However, the historical element to their theory opens them up to much criticism as we have seen. By employing the notion of a general reasons-responsive cognitive power, as I have done above, we can maintain the strength of Fischer and Ravizza's reasons-responsiveness, whilst avoiding the weaknesses of their broad historical and taking responsibility components.

My remixed theory is, like Fischer and Ravizza's, an actual sequence account. In this way, the actual actions that certain mechanisms lead to are most important, indeed, are those which an agent may or may not be held responsible for. In this way, I assume, for the most part, that the mechanisms are necessarily operative, so that such a thing as an executive power, i.e. a power that allows a mechanism to be actually in operation, is present in the mechanisms under scrutiny. One of the main purposes of the use of mechanisms in the remixed theory is to provide something in virtue of which an agent can be judged to be responsible. The other main reason for a mechanism-based theory is Fischer and Ravizza's contention that a focus on the actually operating mechanisms shields the theory from determinism's removing of alternative possibilities, as we saw in (Chapter 2).

In what follows I first outline in greater detail what I have said about the notion of a general reasons-responsive cognitive power (5.1). I then describe further which mechanisms lead to actions that an agent ought to be held responsible for, i.e. which mechanisms are necessary for an agent's responsibility. To do this I will draw on Fischer and Ravizza's historical and taking responsibility notions, but ultimately reduce this down to the authenticity of the relationship between an agent's general

reasons-responsive cognitive power and the operating mechanism (5.2). I also briefly sketch the ways in which one might be held responsible for weak willed behaviour and omissions (5.3), since this is the sort of responsibility that occurs most often in everyday life.

5.1 The General Reasons-Responsive Cognitive Power

The general reasons-responsive cognitive power (GRRCP), as I have briefly outlined, is that which is instantiated in specific instances by the reasons-responsive mechanisms that might lead to an agent's actions. One way to understand this notion is by analogy to another of one's possible general cognitive powers, say one's general philosophical cognitive power. One might characterise an agent's philosophical ability by saying that she has an effective general philosophical cognitive power, meaning that if asked a philosophical question which she has not been asked before, she would be able to answer it using this general cognitive power. The answer she comes up with would be produced by a specific instantiation of her general philosophical cognitive power.

The GRRCP is one that issues in action mechanisms, such as Jones doing x, whereas the general philosophical cognitive power issues in knowledge mechanisms, such as the answers to philosophical questions. The important thing they have in common though is their generality and the subsequent mechanisms that instantiate this general cognitive power. So the GRRCP is a general cognitive power that is instantiated in particular mechanisms that lead to actions.

But how does such a thing develop in an agent? Fischer and Ravizza, as I described in (2.4.2), take the moral education of a child very seriously as part of their taking responsibility condition. Though there are problems with the taking responsibility condition itself, I also think that the way in which a child becomes a

moral agent is important. For the remixed theory, however, a child's moral education, in a very broad sense, is important for the development of their GRRCP. It is not just that the child learns what is considered by her society to be right and wrong, good and bad, but that she understands the reasons why she ought or ought not to do certain things. By treating a child as if she is a moral agent already, one not only teaches her about morality, one also teaches her about acceptable reasons for and against certain actions. Take Fischer's example of the boy at the birthday party ripping open all of the birthday girl's presents in his excitement. In this scenario, the boy not only learns that one ought to control one's excitement and that opening other peoples' presents is wrong, but that there are reasons one might take into account in favour of acting, or indeed not acting, in a certain way. In this example, the parents might show the boy how sad and disappointed the birthday girl is that she didn't get to open her presents herself and he might also learn how unsatisfying giving into excitement that way can be. All of these experiences allow the boy not only to learn how one ought to act in specific situations, but also helps to develop his general reasons-responsiveness. Just as the GRRCP issues specific mechanisms, it can be developed by many specific experiences of trying different mechanisms out and learning from them. What Fischer and Ravizza refer to as a "typical case" (1998, p. 208) of moral education, might also extend just to a typical upbringing in which one can easily learn about reasons.

The reason that the development of the GRRCP is important is also because it is part of the agent's becoming the person she is. One's upbringing shapes one's personality and one's identity and a big part of this identity and personality is the GRRCP. Though it is true that many people will have similar responses to reasons because there is some objective element to what might be considered 'good' reasons, this is by no means uniform. Indeed, one's GRRCP is heavily influenced by the

society in which one is brought up. This is due to the moral standards of societies being different in different places and the meta-ethical difficulty of being able to detect objective moral truths, if indeed there are any. Even within a society there are differences of opinion about what are good reasons for doing things and what are bad reasons. Fischer and Ravizza intend their account of reasons-responsiveness to be theory-neutral with regards to moral philosophy such that they only require that the moral reasons "are at least in the "ball park" as contenders for being correct, [and] are given by the considered judgements (in wide reflective equilibrium) of the relevant community" (1998, p. 77). Taking up this attitude towards moral reasons, there is more room for an agent's GRRCP to be unique to them since reasons-responsiveness does not require an appeal to objective moral truths.[9] On the remixed theory, an agent's GRRCP represents a part of her autonomy and personhood and so is a very important part of the agent as a person.

The remixed theory can also characterise the reasons why some severe forms of indoctrination might remove responsibility for an agent. This is through the non-normal development of the agent's GRRCP. If an agent is indoctrinated with abnormal beliefs, then their responsiveness to reasons may also be abnormal and so their GRRCP will not develop properly if she is a child, or may be warped if she is already an adult. It is difficult to determine whether and to what extent an agent's GRRCP might have been warped, but this represents the difficulties there are in determining whether or not an indoctrinated agent is responsible.[10] The important

[9] For a compatibilist view that does appeal to objective moral truths, see Susan Wolf's reasons view (Wolf, 1990).

[10] Where indoctrination is concerned, one might wish to separate responsibility from blameworthiness such that an agent may be considered responsible for their actions despite their indoctrination, but might not be blameworthy because of their indoctrination.

thing to gain from this is that, should an agent's GRRCP be warped but still maintain something that might resemble reasons-responsiveness, the agent might not be held responsible for those actions issued by mechanisms that are instantiations of this warped GRRCP. I will return to this in (5.2).

We are now in a position to loosely define the GRRCP as a cognitive power that allows an agent to issue reasons-responsive mechanisms for action in specific situations. It is in virtue of the agent's GRRCP that one might call an agent reasons-responsive, and it is in virtue of the reasons-responsive mechanisms that instantiate the GRRCP that an agent might be called responsible, as I described in (3.1).

5.2 Being Responsible For the Action Issued From a Mechanism

The remixed theory I am proposing relies heavily on Fischer and Ravizza's description of reasons-responsiveness and use of mechanisms. It does not, on the other hand, embrace Fischer and Ravizza's account of the ownership of a mechanism. Instead I propose a description of which mechanisms are necessary for an agent's responsibility that incorporates the idea of the general reasons-responsive cognitive power.

On the remixed theory, an agent can be rightly held responsible when the mechanism that issued in her action has an authentic relationship to the agent via the agent's own GRRCP. This authentic relationship is characterised by the instantiation relation between the agent's GRRCP and the reasons-responsive mechanism. It must be the case that the mechanism is an instantiation of the agent's own GRRCP, as opposed to someone else's or an artificial GRRCP, in order for the agent to rightly be held responsible for her action. To see this, take the example of Jones and Black once again. Jones, in the actual scenario, where he does not waver, does x and can be said to be responsible because he acts from a reasons-responsive mechanism that is an

instantiation of his own GRRCP. In the alternative scenario where Black takes over, we might say that Black implants a reasons-responsive mechanism that issues in Jones' doing x, but this mechanism is not an instantiation of Jones' own GRRCP (it might instead be an instantiation of Black's GRRCP, or maybe an artificial GRRCP). In this scenario, even though the mechanism from which Jones acts might be reasons-responsive, Jones is not responsible for his action because he cannot be said to have an authentic relationship to this mechanism since it is not an instantiation of his own GRRCP.

However, this is not to say that a mechanism that comes from elsewhere might not be a *potential* instantiation of an agent's GRRCP. As we saw in (4.3), Fischer's idea that an agent must reflect on an artificially implanted mechanism can be understood as the mechanism being reflected upon through the agent's GRRCP. Through the GRRCP, the agent can determine whether or not the mechanism is acceptable to her, i.e. whether or not it *could* be an instantiation of her GRRCP. In this way, the remixed theory allows for minimal manipulation in the implanting of artificial reasons-responsive mechanisms, but only those that are confirmed as instantiations of the agent's own GRRCP. For example, advertising is a type of manipulation, the consequences of which, an agent would still be responsible for if the mechanism through which she acts is reasons-responsive in virtue of her own GRRCP, *not* in virtue of the artificial GRRCP that the advertisement promotes. I think that this makes the remixed theory quite plausible since there is a strong intuition that someone who has been influenced by advertising has not been manipulated in the same way, or at least not to the same degree, as the type of manipulation that occurs in something like hypnosis, or indeed the manipulation arguments I discuss in (Chapters 6 and 7).

As mentioned above, there is the question of the warping of an agent's GRRCP and how this affects the responsibility the agent has for the actions that issue from the mechanism that is an instantiation of this warped GRRCP. Clearly if the GRRCP has been warped in such a way that it can no longer produce reliable reasons-responsive mechanisms, there is no problem: the agent is not responsible in virtue of her mechanisms not being reasons-responsive. It is when it is warped in such a way that it maintains *a* general reasons-responsiveness, but no longer produces the same sorts of mechanisms it would have originally. In this way it might be that the GRRCP that issues the mechanisms is no longer the original agent's GRRCP, but a new warped GRRCP. As mentioned above, the GRRCP is an important part of an agent's personality and identity, so to change it in this way might produce changes in the agent herself. These changes subsequently affect the plausibility of holding the original agent responsible for these new personality traits. Of course, this is not to say that a person cannot change their personality a little over time, but not to the extent that they suddenly have completely opposite views and character traits to those they possessed before.

Alfred Mele offers an example (Mele, 2008, pp. 266-268), the gist of which nicely demonstrates my point. He imagines two fathers, Pat and Paul. Pat is a good parent and wants to take out a loan to pay for his daughter's university fees. In Mele's terms, he has certain unsheddable values that lead him to this desire. In terms of the remixed theory, we might say that Pat has a certain GRRCP that issues in the sorts of mechanisms that lead him to take out a huge loan to pay for his daughter's university fees. Paul on the other hand, is selfish and does not wish to take out a huge loan to pay for his daughter's university fees. In Mele's terms he has certain selfish unsheddable values. On the remixed theory it might be described that kind, unselfish

mechanisms such as that which leads to Paul taking out a loan to pay for his daughter's university fees, would not be an instantiation of Paul's own GRRCP. Little known to Paul, however, his wealthy mother has hired a team of psychologists to change Paul's GRRCP to something that closer resembles Pat's own GRRCP, such that Paul suddenly has the urge to take out a huge loan to pay for his daughter's university fees. Thus it seems that his personality has changed – he never would have thought of doing such a thing before. Assume that both Pat and Paul do indeed take out loans for their daughters' university fees. It seems that Pat is responsible for his action since the mechanism that led to it is an instantiation of his own GRRCP, whereas Paul is not responsible for his action since such a mechanism is not an instantiation of *his* own GRRCP, it having been warped in such an unnatural way.

So it is that the GRRCP is an essential part of an agent's personality, such that only the mechanisms that are instantiations of the agent's own GRRCP are those relevant for the agent's responsibility. As we have seen with Pat and Paul, when the warping of a GRRCP takes place the situation becomes complex. Though I do not wish to delve into a discussion about personal identity, it seems that these sorts of issues do enter the debate in such scenarios. This, again, is merely supposed to demonstrate that the GRRCP is an essential and important part of an agent as a person. In turn we see that since the GRRCP is so essential to the agent, any mechanism issued from the agent's own GRRCP has an authentic relationship to the agent and thus the agent can be rightly held morally responsible for the actions that issue from this mechanism.

The warping of a GRRCP can be done on a much smaller scale by the agent herself, for example by being intoxicated. The way in which the remixed theory might explain the examples like that of the drunk driver that we saw in (2.4.1), is in virtue of

the temporary and slight warping of the drunk's GRRCP. The mechanism that issues in the drunk driving a car when he has had too much to drink is not reasons-responsive. But the drunk is still responsible because he allowed himself to get too drunk and thus sacrificed the successful action of his GRRCP. So though the mechanism on which the agent acted was not an instantiation of his GRRCP and so not reasons-responsive to the extent that the sober drunk usually is, he is still responsible because it was his own actions, which *did* issue from reasons-responsive mechanisms that led to his compromising his GRRCP. The remixed theory can explain why the drunk driver is responsible for his running over a child in virtue of his own reasons-responsive mechanisms leading to subsequent actions that caused the temporary warping of his GRRCP. This then demonstrates a further feature of the remixed theory: that the agent is responsible for actions that issue from mechanisms that *would have been* informed by the GRRCP, but *could not be* because the agent *herself* had compromised the successful action of her GRRCP.

On the remixed theory, then, an agent is responsible for a mechanism if it is an instantiation of her own general reasons-responsive cognitive power, including when this mechanism leads to an agent knowingly compromising the successful action of her GRRCP.

5.3 Responsibility For Weakness of the Will and Omissions

So far I have been considering actions that issue from mechanisms that are operative, such that the agent is responsible for an action she actually did. We might also hold people responsible, however, for failures of action perhaps from weakness of the will. First I will consider responsibility for omissions and then for weakness of will.

Let us return to my earlier example of Daisy and Amelia: imagine that Amelia has forgotten that it's Daisy's birthday and so fails to get her a present. We would

indeed hold Amelia responsible for her omitting to get Daisy a present, but on what grounds according to the remixed theory? Here we can draw on the same sort of ideas that helped with the responsibility of the drunk driver. Recall that the drunk driver is responsible for his actions whilst drunk because he allowed himself to compromise the successful action of his GRRCP. In the case of Amelia's forgetfulness, we might also invoke the compromising of the successful action of her GRRCP: by forgetting that it was Daisy's birthday, Amelia did not allow her GRRCP to issue in the sorts of mechanisms that would lead her to get a present for Daisy. Amelia is usually a very good friend and had she remembered Daisy's birthday, her GRRCP certainly would have issued in the sorts of mechanisms that led to her getting Daisy a present, probably the bracelet Daisy so desires. But this time, Amelia's forgetfulness compromised the successful action of her GRRCP, so no such mechanism was issued. To the extent that an agent is responsible for their forgetfulness and other such omissions, she is responsible for the subsequent effect this has on the successful action of her GRRCP. Thus, an agent can be responsible for omissions in such a case where she prevents the successful action of her own GRRCP. This relies on the counterfactual that, given the right input, the agent *would have* acted on mechanisms that were an instantiation of her own GRRCP.

The explanation for responsibility for behaviour stemming from weakness of the will is only slightly different. Recall Fischer and Ravizza's description of reasons-responsiveness (2.3), particularly the asymmetry between strong reasons receptivity and weak reasons reactivity. What is important here is the weak reactivity to reasons. Indeed, one recognising reasons to do or not to do something, and yet not being reactive to these reasons can explain weakness of the will. For example, Amelia remembers that it is Daisy's birthday and intends to get her a present. However,

Amelia cannot be bothered to go to the further shop where Daisy's desired bracelet can be bought and so does not get her the bracelet she really wants. Amelia recognises the reasons she has to go to the further shop to get the bracelet for Daisy, but nonetheless is not reactive to these reasons owing to her laziness, i.e. her weakness of will. This failure to react to reasons can be characterised by Amelia again preventing the successful action of her GRRCP by being lazy. To the extent that an agent is responsible for weakness of the will, she is responsible for the subsequent effect this has on the successful action of her GRRCP, particularly on the agent's reactivity to reasons.

As is clear, the successful action of the GRRCP is an essential element of the remixed theory. It is not merely that the agent's mechanisms be instantiations of the agent's own GRRCP, but that the GRRCP must be working properly, i.e. not warped and not compromised by something like alcohol, forgetfulness or weakness of will.

Here ends the description of the remixed theory of responsibility. In the following two chapters I will examine two arguments pitted against any compatibilist theory of moral responsibility, which I wish to address in Chapter 8. The first (Chapter 6) is Pereboom's four case manipulation argument, and the second (Chapter 7) is Mele's zygote argument.

6. Pereboom's Four Case Manipulation Argument

In this chapter I examine Pereboom's argument, first explaining the argument itself (6.1), then showing the ways in which Fischer and Ravizza's theory is threatened by it and their replies (6.2) and finally considering another reply to Pereboom's argument, namely that from Michael McKenna (2008) (6.3).

6.1 The Argument

In his (2001) book *Living Without Free Will*[11] Pereboom sets out his Four Case Manipulation argument which aims to undermine compatibilist views about moral responsibility and support his version of determinism. Here I focus on Pereboom's attack on Fischer and Ravizza's theory, though Pereboom intends his argument to apply to any compatibilist requirements for freedom, notably, Frankfurt's theory, which concerns the harmony between first and second order desires.

Pereboom's Four Cases may be paraphrased as follows (2001, pp. 112-115):

> Case 1: Our agent, Professor Plum, is like an ordinary human being, but he was created by neuroscientists and can be manipulated directly by them, say, using remote controls. The neuroscientists manipulate Plum, such that his reasoning processes are reasons-responsive, to kill Ms White.
>
> Case 2: Our agent, Plum, is like an ordinary human being, but he was created by neuroscientists, who have programmed him such that his reasoning processes are reasons-responsive and lead him to kill Ms White.
>
> Case 3: Our agent, Plum, is an ordinary human being, but he is determined by the rigorous training practices of his home and community such that his reasoning processes are reasons-responsive and lead him to kill Ms White. The training happened too early in his life for him to have been in control of this determination.

[11] The book section is adapted from (Pereboom, Determinism Al Dente, 1995)

> Case 4: Physical Determinism is true. Our agent, Plum, is an ordinary human being raised in ordinary circumstances. Plum kills Ms White as a result of his reasons-responsive reasoning process.

According to Pereboom, the steps from case 1 to 2, 2 to 3, etc. do not involve any significant change in the moral requirements for responsibility, such that if you accept that the agent, Professor Plum, cannot be held morally responsible in case 1, nor can he be held morally responsible in cases 2, 3, or 4. Pereboom ultimately concludes that since determinism is true and there is no difference between each of these cases, then we cannot hold anyone morally responsible.

6.2 Fischer and Ravizza vs. Pereboom

Pereboom intends to show that Fischer and Ravizza's responsibility criterion of reasons-responsiveness is not in fact sufficient for an agent to be held morally responsible and that their theory does not block determinism as they intended. An obvious rebuttal that might be made by proponents of Fischer and Ravizza's theory is to appeal to the historical component. As I described above, Fischer and Ravizza intend that such cases as Pereboom's case 1 are ones in which the agent is not held morally responsible because of the artificial historical connection between the agent and the mechanism on which the agent acts. Indeed Fischer (2004) contends that there is no impediment to holding Plum responsible in case 1, and so it is not a problem that this conclusion could be taken ahead to case 4, where one would also hold Plum responsible.

Fischer contends that it is not just that Plum is responsible in case 1 because of the historical component, but that reasons-responsiveness is not something that can be artificially implanted in an agent: "The reasons-responsiveness itself cannot have been put in place in ways that bypass or supercede the agent - the mechanisms that issue in one's behavior must be one's own" (2004, p. 147). Indeed, this reply of

Fischer's can be even better characterised by the remixed theory as we shall see in (8.1).

Also in his (2004) paper, Fischer points back to his and Ravizza's contention that "certain cases of significant manipulation that occur literally from birth (or in this case from the beginning of the existence of Professor Plum) there is no opportunity for a self to develop" (2004, p. 156). If this contention were to be honoured, it would seem as though case 2 was not sufficiently similar to the other cases because here, a real self does not develop.

Fischer further makes use of an important distinction he holds between moral responsibility and blameworthiness. He says that "moral responsibility, as Ravizza and I understand the notion, is more abstract than praiseworthiness or blameworthiness: moral responsibility is, as it were, the "gateway" to moral praiseworthiness, blameworthiness, resentment, indignation, respect, gratitude, and so forth" (2004, p. 157). In this way, even though one might hold Plum morally responsible in case 1 and thus also, according to Pereboom, in cases 2, 3, and 4, one might not *blame* him owing to his manipulation. Fischer claims that though there is no difference between the four cases in terms of responsibility, there is in terms of blameworthiness. He says that this is "a function of the circumstances of the creation of his values, character, desires, and so forth" (2004, p. 158). In case 4, there does not seem to be any reason to suppose that the creation of Plum's values etc. are unusual and therefore there is reason to think Plum blameworthy for the death of Ms White in case 4, even if this is not so in case 1. This idea of the normal creation of values etc. being important for responsibility is something that the remixed theory characterises through the development and importance of the GRRCP, as we have seen. So this line

of defence against Pereboom's argument is also open to the remixed theory, as I will further outline in (8.1).

6.3 Other Replies to Pereboom's Argument

Fischer's argument that one might reasonably hold Plum responsible in case 1 and thus not have to concede Pereboom's conclusion that determinism removes moral responsibility, is in line with the compatibilist intuition about case 4. Michael McKenna (2008) argues that a "hard-line" response to Pereboom's argument can run with the compatibilist intuition about case 4 to forward their agenda, just as Pereboom runs with the intuition that Plum is not responsible in case 1 to forward his *in*compatibilist agenda. A compatibilist can take case 4 as one in which Plum *is* morally responsible for killing Ms White, and then by virtue of Pereboom's making the cases so similar, Plum in case 1 is not morally responsible either. In fact, McKenna argues that since it is not clear, if we start with Pereboom's case 4, whether or not Plum is responsible in this case, this uncertainty is then carried right through to case 1. This means that the intuition Pereboom intends for case 1, i.e. that Plum is not morally responsible, is not as obvious as he supposes (McKenna, 2008, p. 153). This in turn leaves the debate at a stalemate that simply comes down to the different intuitions of each side: the compatibilist verses the incompatibilist.

McKenna argues that given that it was the incompatibilist side that proposed this argument, the burden is on the incompatibilist, not the compatibilist to make the next move. McKenna's position is that in this scenario, the compatibilist need only defend their claims against the incompatibilist arguments, such as Pereboom's, rather than positively give arguments in favour of compatibilist intuitions (2008, p. 148). Pereboom, on the other hand, does not see it this way as McKenna describes:

> While acknowledging that some might not share his intuitions, Pereboom holds fast to his incompatibilist convictions regarding Plum. He denies that we wind up here with a dialectical stalemate. He holds that the intuitive scales are tipped in his favor. To this, I can only voice my disagreement with him. I think that our intuitions do not clearly speak in Pereboom's favor. If I am correct, if this disagreement does end in a stalemate, then this amounts to a victory for the compatibilist, since she was only out to defeat an argument for incompatibilism, not to prove her compatibilist thesis.
>
> (McKenna, 2008, p. 154)

It seems that, whichever side the burden is on, there is a stalemate in this debate that simply comes down to whether one has a compatibilist or an incompatibilist intuition in the first place. This certainly seems to weaken the force of Pereboom's four case argument against compatibilism. McKenna also adds that there are in fact many instances of "mundane manipulation" that occur in everyday life, which we do not take to be responsibility undermining. In fact he suggests that they might support the idea that we do in fact have free will. McKenna describes a woman who has been shaped by her childhood experience of her mother's long and painful death through a fight with leukaemia, to be someone who has certain values about the preciousness of life and how one must make the most of it. This turn of events looks like a real life manipulation case, but in fact McKenna argues that these sorts of circumstances are what demonstrate one's agency. As McKenna sees it, this circumstance "surely does not undermine her free and responsible agency. It makes it" (2008, p. 156). The idea seems to be that there are many ways in which one could have reacted to the early death of a parent, and this particular woman chose the values that she took from that experience. McKenna concludes that:

> Unfortunately, in my estimation, all these cases can do is soften one who entertains the Manipulation Argument to the mere possibility that dramatic full-blown science fiction manipulation cases need not clearly be freedom and responsibility undermining. They will not be a proper basis for moving one to the further conclusion that cases like Pereboom's are clearly not freedom and responsibility undermining.
>
> (2008, p. 157)

In other words, McKenna rejects the force of Pereboom-type manipulation arguments because the intuitions they might invoke are not true to life, the situations being so alien.

I think that McKenna's reply to Pereboom's argument is somewhat unsatisfying. Compatibilists do seem to be able to at least positively support their theses by presenting plausible arguments like Frankfurt's against the principle of alternative possibilities described above (1.1). Furthermore, it is the nature of philosophical thought experiments that they take liberties with reality as we know it. The existence of these alien situations, which these thought experiments describe, does not mean that we should not *at least try* to test our intuitions on them.[12]

[12] There is also the question of whether or not similarity is transitive in the way that both Pereboom and McKenna assume it is: does A being similar to B and B being similar to C necessarily imply that A is similar to C? I have not discussed this issue here, but it is something worth keeping in mind in this dialectic.

7. Mele's Zygote Argument

Alfred Mele (2008) critiques Pereboom's four case argument on the grounds that his "best-explanation premise" for the intuition that Professor Plum is not responsible in the first three cases relies on the determinism in the examples, such that those that have this intuition might not have it if indeterminism were present. Let us examine this a little closer.

Mele argues that to test the inference to the best explanation that Pereboom appeals to, one must separate the manipulation in the cases from the determinism in the cases. When this is done, "we should expect intuitive incompatibilists to have incompatibilist intuitions and intuitive compatibilists to have compatibilist intuitions" (2008, p. 277). If this is so, then even in Mele's proposed indeterministic stories, incompatibilists have the intuition that Professor Plum is not responsible. But the best explanation for this intuition clearly cannot be that Plum's "actions result from a deterministic causal process that traces back to factors beyond his control" (Pereboom, 2001, p. 116), since in the indeterministic cases that Mele proposes, the processes are not deterministic. Mele concludes that the non-responsibility intuition must come out of the manipulation, not the determinism, in the four cases. Thus, Pereboom's best-explanation premise does not seem to be plausible. Since Pereboom's four case argument relies on the best-explanation premise, there needs to be a strong argument for the plausibility of the best explanation that persuades compatibilists. Mele does not think that Pereboom does this and so offers his own argument to remedy this issue.

It is this argument, namely Mele's zygote argument, to which we now turn. I will first outline the argument itself (7.1), then describe how it threatens Fischer and

Ravizza's theory and discuss some possible replies of theirs (7.2) and finally outline a further reply to the argument itself from Stephen Kearns (2012) (7.3).

7.1 The Argument

Mele, in his (2008) paper, sets out an argument that is supposed to show that the compatibilist conditions for responsibility (whatever they might be) are not sufficient to hold an agent responsible, much like Pereboom's manipulation argument.[13] Mele imagines a scenario in which Diana creates a zygote in Mary that will produce an agent, Ernie, who fulfils the relevant compatibilist conditions, for our purposes an agent who acts on reasons-responsive mechanisms. Diana programmes the zygote such that Ernie will do some action A at a specific time t, say 30 years from now. Ernie seems to satisfy the relevant compatibilist conditions for responsibility for his action A, but can we still say he is responsible? Mele then suggests that this agent Ernie is really no different from Bernie, another agent very similar to Ernie, but whose zygote came about in the normal way. What difference does the zygote make to the responsibility of each agent? Mele formalises his argument as follows:

1. Because of the way his zygote was produced in his deterministic universe, Ernie is not a free agent and is not morally responsible.
2. Concerning free action and moral responsibility of the beings into whom the zygotes develop, there is no significant difference between the way Ernie's zygote comes to exist and the way any normal zygote comes to exist in a deterministic universe.
3. So determinism precludes free action and moral responsibility.

(Mele, 2008, p. 280)

[13] It is contested whether or not Mele's zygote argument is a manipulation argument.

The first premise characterises the idea that Diana's intervention and programming in the production of Ernie's zygote somehow impacts on Ernie's subsequent responsibility. The second premise, the "no-difference premise", asserts that there is no difference between Ernie and Bernie, or more generally an agent from a manipulated zygote and one from a normal zygote. Assuming that both of these premises are true, it does indeed follow that determinism precludes moral responsibility.

7.2 Fischer and Ravizza vs. Mele

Fischer and Ravizza's theory of moral responsibility looks to be challenged here particularly because it seems that an appeal to the historical notion they employ might not be appropriate. To say that the difference between Ernie and Bernie is that Ernie's zygote was artificially produced stretches their notion of ownership of a mechanism a little far. Another way in which one might defend compatibilism against Mele's argument is by appeal to something similar to what Fischer said about personhood, namely, that Ernie is in fact not a person since he did not have the chance to become one, his zygote having been artificially produced and pre-programmed by Diana.

In fact, Fischer directly replies to the zygote argument (2011), claiming that it does not pose a threat to compatibilism. This, he argues, is due to the fact that the first premise need not be accepted since "… the distal intentions of the agents who bring Ernie into being – [are] irrelevant to Ernie's moral responsibility when he matures into an adult many years later" (2011, p. 268). Fischer motivates this claim by comparing Mele's case of Diana creating the zygote and an alternative one where Ernie's parents intend to make a zygote which turns out to be Ernie. He admits that in the Diana case, one's intuition is perhaps different to that in the normal parent case, but there is no reason for this. Indeed, it seems that "the basis for Ernie's

responsibility is more 'local'" than something as remote as the zygote production (2011, p. 268). Fischer does not elaborate here on what exactly he means by "local", but I will argue in 8.2.1 that "local" can be characterised using the remixed reasons-responsive theory.

Fischer further suggests that Mele's second premise, the no difference premise, is something that, on its own, compatibilists can embrace. To show this he imagines a case in which Mary is in a clinic during the night and if a random number generator selects 1, then a zygote will be placed into her, but won't be if the random number generator selects 2. Since the random number generator selects 1, Mary gets the zygote (2011, p. 271). Fischer sees no relevant difference between this way of implanting the zygote and the case in which Diana implants the zygote. In both scenarios it seems not to make a difference to Ernie's subsequent responsibility. Thus Fischer rejects that Mele's zygote argument is a problem for his, or any other, compatibilist view. It is important to see that Fischer rejects the argument ultimately by rejecting that Ernie is *not* responsible for his actions. In this way Fischer does not confront the possible intuition that Ernie is not responsible for his actions, which then assumes that the zygote case *is* sufficiently similar to the incompatibilist's opinions about Determinism. In 8.2.2 I will show how one might draw a distinction between Mele's zygote scenario and the regular deterministic scenario.

7.3 Other Replies to Mele's Zygote Argument

Rather than arguing with the premises of Mele's zygote argument, one might attack its validity as Stephen Kearns (2012) does. Kearns suggests that the meaning of premise 1 is not entirely clear. By "because of the way his zygote was produced in his deterministic universe, Ernie is not a free agent and is not morally responsible" does Mele mean either that (a) Ernie's actions are deterministically caused by the zygote,

(b) Ernie's actions were manipulated, (c) Ernie's actions were deterministically manipulated, or a disjunction of the three (2012, pp. 381-382)? If we take meaning (a) in premise 1, it looks something like this: "Because the structure of his zygote and all of his actions were deterministically caused, Ernie is not a free agent and is not morally responsible for anything" (Kearns, 2012, p. 382). If this is the way that Mele intends premise 1, then it is clearly question begging since only an incompatibilist would accept this premise. Kearns holds that a similar claim works for (c) as well:

> If the combination of manipulation and determinism renders Ernie not free, then determinism alone does not render agents not free, meaning that there is a significant difference between Ernie's case and the case of normal agents in deterministic worlds. If, on the other hand, a defender of the argument says that determinism alone is sufficient to render Ernie unfree, then she must deny 1c. Ernie lacks freedom simply because his actions are deterministically caused, not because they are deterministically manipulated. Given this, she cannot appeal to 1c in an argument for incompatibilism (as it is false even by her own lights). Either way, the zygote argument (with premise 1 interpreted as 1c) fails.
>
> (2012, p. 384)

Furthermore, Kearns argues that premise 1 interpreted with meaning (b), such that Ernie's actions are manipulated, only holds traction in a deterministic universe. To see this, Kearns invites us to consider the Ernie scenario in an indeterministic universe. In this case, it doesn't seem like the no difference premise holds; manipulation is not relevantly similar to determinism. It then looks as though Mele relies heavily on the fact of the Ernie case being in a deterministic universe, which shifts the focus from the manipulation to the determinism (2012, p. 386), which, as we have seen, Mele criticises Pereboom for.

Finally, if one takes Mele to mean a disjunction of (a), (b), and (c), then the premise looks difficult to defend. Clearly it can't be defended on the basis of any of the individual meanings since these are all problematic, as we have seen. The only thing left is an appeal to intuition (2012, p. 387). The problem with this is that the

compatibilist and the incompatibilist have opposite intuitions and so will never agree on this particular premise, based on intuition. Thus Kearns, similarly to Fischer, argues that Mele's zygote argument is not something that compatibilists need take a stand on since "if deterministically manipulated agents are unfree, then there is no non-question-begging reason to believe that this lack of freedom transfers to normal determined agents (and, of course, if deterministically manipulated agents are free, then incompatibilism is straightforwardly false)" (2012, pp. 388-389).

Between them, Fischer and Kearns have made the threat of Mele's zygote argument look far less daunting. There is, however, still the possible intuition that Ernie is *not* responsible *because of* Diana's having programmed his zygote. If one accepts this, one then has to explain why this scenario is any different from the regular deterministic scenario, which, as I have said, I will try to do in the remixed theory's compatibilist terms in (8.2).

8. The Remixed Theory and Manipulation Arguments

8.1 Reply to Pereboom's Argument

Aside from the replies to Pereboom's argument that I have outlined from Fischer and McKenna, there are still some things that the remixed theory can add to these to make the compatibilist side stronger.

In the original version of Pereboom's Four Case Manipulation argument (Determinism Al Dente, 1995), he suggests that Fischer might deny that the agent could act on a reasons-responsive mechanism through direct stimulation of the brain. This is due to Fischer's remark that "in a case of direct manipulation of the brain, it is likely that the process issuing in the action is not reasons-responsive, whereas the fact that a process is causally deterministic does not in itself bear on whether it is reasons-responsive" (Fischer, 1987).[14] Pereboom flatly denies this saying that "as long as a process requires only abilities that are physically realised, it can be induced by sufficiently equipped scientists" (Pereboom, 1995, p. 24). This discussion then seems to have been abandoned, but I think there is something important we can take from it. It seems to be that reasons-responsiveness must necessarily be produced by an agent herself, not merely through physical stimulation. One might argue that, however advanced brain science might get, it is in fact a category error to try to reduce reasons-responsiveness to a physical process in the brain. Indeed, when we talk about mechanisms here and in other areas of moral philosophy, rarely do we mean the physical and chemical processes that might occur in the brain when an agent acts. Instead we use the notion of a mechanism to "point to" the macro process that goes on

[14] Reprinted in (Fischer, My Way: Essays on Moral Responsibility, 2006).

in the mind. I do not wish to get into a deep discussion about the mind/body problem and consciousness here, needless to say, it seems more plausible that the notion of reasons-responsiveness is not something that can be artificially induced through direct stimulation of the brain, since, as I mentioned before, it seems to be an irreducible macro-property. Nonetheless, there is a better way to characterise Fischer's intuition that direct stimulation of the brain cannot produce reasons-responsive mechanisms, namely through the remixed reasons-responsive theory.

The idea that Fischer gestures towards links to what I have said previously about the necessity that mechanisms be instantiations of the agent's own general reasons-responsive cognitive power. In the case of direct stimulation of the brain, as in Pereboom's case 1, the mechanism that issues the agent's action is certainly not a direct instantiation of the agent's own GRRCP, and it seems likely (though not certain) that the mechanism has not even been analysed as a potential instantiation of the agent's GRRCP. In this way, we can characterise Professor Plum's lack of responsibility for his action in case 1 as an inauthentic mechanism.

Similarly, in case 2, Professor Plum's mechanism is inauthentic because it is not an instantiation of Professor Plum's own GRRCP, but instead is an instantiation of the pre-programmed GRRCP that the evil neuroscientist created. Professor Plum is created by the neuroscientist and presumably does not have a typical moral education. Without a typical moral education, an effective GRRCP cannot develop and so it seems unlikely that the mechanism that issues in Plum's murdering Ms White is not reasons-responsive. However, this might not be a charitable reading of Pereboom's argument and so we shall assume that Plum did have a typical moral education and can be considered to have developed an effective GRRCP. In this instance, Pereboom's case 2 looks a lot like Mele's zygote case with Ernie and Diana. Just as

Ernie's zygote is created and programmed by Diana, we might assume that Plum's zygote is created and programmed by the neuroscientist. But as we shall see in (8.2), there is a way, on the remixed theory, to show why Diana creating Ernie's zygote, or the neuroscientist creating Professor Plum, has an effect on the responsibility of the agent.

When we then move to case 4,[15] the remixed theory does not accommodate saying that Plum is not responsible for killing Ms White. Professor Plum fulfils all criteria for responsibility as set out in the remixed theory. This is because, it being a compatibilist view, the remixed theory allows for responsibility in deterministic worlds, indeed that is how it is designed. So it seems that there is in fact a relevant difference between the cases. This reply to Pereboom's argument essentially says that he has not captured the compatibilist conditions sufficiently, something that McKenna says is useless. However, I argue that it is not just that Pereboom has not captured the conditions of responsibility, but that it is not possible to do so in the kinds of cases he wants to present. This is due to the fact that the conditions in the remixed theory and the type of manipulation Pereboom proposes are incommensurable. So it is just not open to Pereboom to make a four case argument that will capture all the conditions of the remixed theory. This reply is stronger than accepting the stalemate that McKenna suggests and is no more based on intuition than any compatibilist argument is.

[15] I have skipped case 3 since the intuition for this case is less clear and it doesn't make so much difference to the whole argument. The remixed theory would probably rule out calling Plum responsible in this case since it is like an indoctrination case; he hasn't had a fair chance to develop an effective GRRCP.

8.2 Reply to Mele's Argument

The GRRCP can also help us with Mele's zygote argument. Though I have shown through both Fischer and Kearns' replies to Mele's argument that it is not as threatening as one might have originally conceived, it might still worry some compatibilists. In this section I present two ways in which the remixed theory of moral responsibility both adds to Fischer's reply to the zygote argument by providing a way to characterise "local" (8.2.1) and a possible way to tackle the intuition one might have that Ernie is in fact not responsible for his action in Mele's original Diana case (8.2.2).

8.2.1 Characterising "local"

Part of Fischer's reply to Mele's zygote argument (as presented in 7.2) is to say that "the basis for Ernie's responsibility is more 'local'" (Fischer, 2011, p. 268). However, Fischer himself does not offer a way of explaining how one might characterise "local" in his framework. Indeed, Fischer and Ravizza's theory has these historical elements which do suggest, as we have seen, that one might trace back as far as necessary. In the case of Ernie, it does seem like there is something in Ernie's performing action A that can be traced back to Diana and her creation of the zygote. Needless to say, this is precisely why Watson and others think it dangerous to introduce an historical component into a compatibilist theory.

The remixed theory, on the other hand, explains the relationship between an agent and a mechanism in a more contained way and thus can characterise "local" more efficiently. On the remixed theory, Ernie's responsibility can be characterised by his acting from reasons-responsive mechanisms that are instantiations of his own GRRCP, which has been typically developed. Since it is the action A that we are looking at and his current GRRCP, the responsibility Ernie has is in this sense local:

Ernie's responsibility stems from his GRRCP and the specific instantiation of that in the reasons-responsive mechanism that issues in his doing A. In this way, assuming that his GRRCP has been typically developed, Ernie is responsible for his actions since he fulfils all responsibility criteria according to the remixed theory. The possibility of Ernie's GRRCP *not* being developed in a normal way is an issue that I discuss next.

8.2.2 Why Might Ernie Not Be Responsible?

As we have seen, Mele intends the intuition about Ernie to be that he is not responsible for his action because of Diana's having created and programmed his zygote. In the preceding defence against Mele's argument, I denied this intuition and argued that Ernie *was* in fact responsible despite his zygote being programmed, similarly to Fischer's reply. The reason for this intuition is not just because it allows me to deny Mele's argument and maintain a compatibilist view, but also because it is difficult to see how such a situation really does remove responsibility.

One way in which proponents of the remixed theory might defend against Mele's zygote argument is to say that the GRRCP could not be developed in the normal way in order for Diana to be successful in her programming. Diana would either have to decide each step of Ernie's whole life leading up to his action A, or she would have to implement some device that causes Ernie to do A at the point Diana wants him to. In the former case, it is pretty clear that Ernie's GRRCP has not been able to develop in a typical way, if at all, and so we would not hold Ernie morally responsible. In the latter case, it seems as though Ernie's GRRCP is allowed to develop in a typical way, but then the mechanism that issues in his doing A is not an instantiation of his own GRRCP. Instead it overrides whatever other mechanism Ernie's own GRRCP might have issued at that moment. In either case, Ernie is not

responsible for his doing A. Since there doesn't seem to be another way to characterise how Diana might programme Ernie's zygote, it seems that we cannot hold Ernie morally responsible yet Mele's intended conclusion fails. This is due to the fact that Diana programming Ernie's zygote is not sufficiently similar to the deterministic world.

This reply essentially contradicts Mele's assumption that Diana's production of the zygote does not involve manipulation. I argue here, that the pre-programming necessary for the zygote argument scenario to go forward is a form of manipulation, as I have described above. In this way, I intend to have shown that Mele's assumption, that his zygote argument avoids manipulation, is wrong.

Concluding Remarks

Whether or not determinism is true, the desire to hold people morally responsible for their actions persists. Ever since Frankfurt's demonstration that one does not need alternative possibilities in order to be held morally responsible, it seems more plausible that even if determinism is true, people can still be held morally responsible. The theories that have been borne out of Frankfurt cases, such as Fischer and Ravizza's, of course have their own problems. What I hope to have shown here is that all is not lost despite these problems and many of Fischer and Ravizza's ideas can be maintained in the remixed theory. The addition of the general reasons-responsive cognitive power to the main idea of reasons-responsiveness as spelled out by Fischer and Ravizza allows for some of the merits of theories such as Smith's and Wallace's, whilst avoiding the problems they incur. Furthermore, the remixed theory is able to properly tackle the strongest arguments against compatibilist views, namely those from Pereboom and Mele, as well as describe more standard cases such as the responsibility of the drunk driver and responsibility for omissions and weakness of the will. By amalgamating the specificity of mechanism-based theories with the generality of cognitive power type theories, the remixed theory provides the 'best of both worlds', as a good remix should.

Reference List

Fischer, J. M. (1987). Responsiveness and Moral Responsibility. In F. Schoeman (Ed.), *Responsibility, Character, and the Emotions.* Cambridge University Press.

Fischer, J. M. (1996, July). Review of Wallace, R. Jay Responsibility and the Moral Sentiments. *Ethics*, 850-853.

Fischer, J. M. (2000). Chicken Soup for the Semi-Compatibilist Soul: Replies to Haji and Kane. *The Journal of Ethics, 4*, 404-407.

Fischer, J. M. (2000). Excerpts from John Martin Fischer's Discussion with Members of the Audience. *The Journal of Ethics* (pp. 408-417). Kluwer.

Fischer, J. M. (2000). Responsibility, History and Manipulation. *The Journal of Ethics, 4*, 385-391.

Fischer, J. M. (2004). Responsibility and Manipulation. *Journal of Ethics*, 145-177.

Fischer, J. M. (2006). *My Way: Essays on Moral Responsibility.* Oxford University Press.

Fischer, J. M. (2011). The Zygote Argument Remixed. *Analysis, 71*, 267–272.

Fischer, J. M. (2012). *Deep Control.* Oxford University Press.

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

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

Haji, I. (2000). On Responsibility, History and Taking Responsibility (Comments on John Martin Fischer's Presentation). *The Journal of Ethics, 4*, 392-400.

Kane, R. (2000). Non-Constraining Control and the Threat of Social Conditioning (Comments on John Martin Fischer's Presentation). *The Journal of Ethics, 4*, 401-403.

Kearns, S. (2012). Aborting the Zygote Argument. *Philosophical Studies, 160*, 379-386.

Lenman, J. (2011, Winter). *Reasons for Action: Justification vs. Explanation.* Retrieved March 2016 from Stanford Encyclopaedia of Philosophy: http://plato.stanford.edu/entries/reasons-just-vs-expl/

Long, T. R. (2004). Moderate Reasons-Responsiveness, Moral Responsibility and

Manipulation. In J. K. Campbell, M. O'Rourke, & D. Shier (Eds.), *Freedom and Determinism.* Cambridge: MIT Press.

McKenna, M. (2001). Review of Responsibility and Control: A Theory of Moral Responsibility. *The Journal of Philosophy, 98* (2), 93-100.

McKenna, M. (2008). A Hard-Line Reply to Pereboom's Four Case Manipulation Argument. *Philosophy and Phenomenological Research*.

Mele, A. R. (2008). Manipulation, Compatibilism and Moral Responsibility. *The Journal of Ethics , 12* (3/4), 263-286.

Pereboom, D. (1995). Determinism Al Dente. *Nous, 29* (1), 21-45.

Pereboom, D. (2001). *Living Without Free Will.* Cambridge University Press.

Smith, M. (2003). Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion. In S. S. Tappolet, *Weakness of Will and Practical Irrationality.* Oxford University Press.

Wallace, R. J. (1996). *Responsibility and the Moral Sentiments.* Harvard University Press.

Watson, G. (2004). Responsibility and the Limits of Evil. In G. Watson, *Agency and Answerability: Selected Essays* (pp. 219-259). Oxford University Press.

Wolf, S. (1990). *Freedom Within Reason.* Oxford University Press.