

Institution of Philosophy, Linguistics and Scientific Theory
University of Gothenburg
Fall 2012

Assessing Historical Compatibilism

Author: Juho Koistinen

Supervisor: Ragnar Francén Olinder

1. The Problem	2
2. Moral Responsibility as Reasons-Responsiveness	3
2.1. Frankfurt Examples	4
2.2. The Methodology	6
2.3. Moderate Reasons-Responsiveness	8
2.4. Taking Responsibility	12
3. History-Sensitive Compatibilism	16
3.1. A Critique of Fischer and Ravizza	16
3.2. Conditions for Free Agency	19
4. Concluding Remarks	22
References	25

1. The Problem

It is a widely shared view that free agency is a requisite for morally responsible action. Acting, say, under duress or drugged does not capture the idea of free action and therefore agents under these circumstances are rarely held morally responsible. Thus, we can conclude that when we are talking about morally responsible agents, we are talking about free agents as well. However, is it conceivable that there are agents acting at least sometimes freely and morally responsibly? This question arises if we consider the possible truth of causal determinism, which is a doctrine that, roughly, says that every event is caused by an event prior to it. Hence, according to causal determinism, every event is necessitated. That is, if causal determinism obtains, every event is taking place necessarily, caused by an event prior to it that is in turn caused by an event prior to it and so on. (Fischer & Ravizza 1998, 14.) It follows from this causal chain that if causal determinism obtains, then there is only one possible future. This calls into question our conception of free will, for if one's efforts leading up to an act do not manifest one's own free will, is there room for free and morally responsible agency or are we just buffeted by causal forces? Those who hold to the latter, incompatibilists, can argue as follows. (1) If a person acts of her own free will, she must have power over her actions. (2) If causal determinism is true, then only one future is possible, given the actual past and the laws of nature. (3) One is powerless over the facts of the past and the laws of nature. (4) Thus, if causal determinism obtains, one is powerless over one's actions. (Ibid., 20—1.) This indirect argument for incompatibilism entails the following two plausible assumptions: one is powerless over the laws of nature and the facts of the past. Therefore, if one is powerless over the facts of the past and if one is powerless over the fact that, given the past, the laws of nature involve a unique future, then one is powerless over the facts of the future (McKenna, 2009).

It is unclear if this is the case, but it is not a philosopher's task to show if causal determinism obtains. Rather, we need to operate under the possibility of the truth of causal determinism and develop a notion of free will and moral responsibility that takes into consideration the possibility of causal determinism. Therefore, we are interested in what the first premise of the indirect argument for incompatibilism refers to. In a normal deliberation, we think that we indeed are capable of altering the future in the direction we think is the best. As for this capability, it is usually thought that we can choose among the alternatives that are in a branching fashion open to us. We are thought to be free when we are able to choose (and subsequently do) otherwise. To put it metaphorically, if freedom of will is thought to be the freedom to choose among the paths we encounter along our lives, then, what is jeopardized by determinism is the ability to choose the other

paths. Thus, determinism would rule out the possibility to do otherwise. (Fischer & Ravizza 1998, 20.) Recall that moral responsibility is closely associated to free agency and then we are ready to state the main problem that this paper concerns. If determinism obtains and it would rule out the freedom to do otherwise, it seems that determinism would rule out moral responsibility as well.

As noted above, incompatibilists endorse the standpoint that the freedom relevant to moral responsibility is not compatible with determinism. The opposite camp, compatibilists, holds that the freedom to do otherwise and moral responsibility are compatible with determinism. Whereas incompatibilists content themselves with arguing that there is no such a thing as moral responsibility to talk about, for a compatibilist the task is a bit more challenging.¹ She needs to provide an account of moral responsibility that is not only consistent with determinism but also captures our intuitions about moral responsibility. John Martin Fischer and Mark Ravizza meet these challenges head on in their book “Responsibility and Control: A Theory of Moral Responsibility” (1998). Their strategy proceeds to show first that the freedom relevant to moral responsibility is not actually the freedom to do otherwise, after which they address the problem of providing an adequate account of moral responsibility. I will then move on to Alfred Mele’s critique of Fischer and Ravizza from the book “Free Will and Luck” (2006). Eventually, I will bring out his alternative compatibilist proposal.

The aim in this paper is to shed light on the problems that describing compatibilist-friendly moral responsibility invokes and, in my opinion, determinism works here as a kind of a frame story. That is, determinism poses a threat to our conception of moral responsibility, and if it is argued that there is no such a thing as moral responsibility, then there is nothing to talk about. However, when we give a chance to moral responsibility, we can carefully develop a notion that allows room for free agency and moral responsibility upon which our deepest intuitions of human life (e.g. chance to meaningful moral relations) are based. These most fundamental intuitions need not be at odds with determinism. I will thus argue for compatibilism and, in particular, for the historical approach of it.

2. Moral Responsibility as Reasons-Responsiveness

The main concept Fischer and Ravizza work with is “guidance control.” Moral responsibility is based on guidance control that is separated into two components. First, in order for one to be morally responsible, a mechanism that issues in an action needs to be responsive to reasons. However, this is not enough but, secondly, an agent’s reasons-responsive mechanism needs to be her own. Before these issues, it is in order to address the challenge that moral responsibility faces

¹ Libertarians are incompatibilists arguing that causal determinism and free will are incompatible as free will rules out determinism. Therefore, the incompatibilist can argue for moral responsibility as well, but in this paper, libertarians are left aside, and by an incompatibilist I will refer to those arguing that determinism rules out moral responsibility.

from determinism.

2.1. Frankfurt Examples

Fischer and Ravizza introduce two sorts of control, guidance and regulative control. Regulative control involves a “dual power” over an action. That is, in order for one to exercise regulative control, one must be able to freely to do some action X and freely to do something else, including refraining from doing anything. (Ibid., 31.) Ravizza and Fischer invite us to consider the following example. Sally drives her car, which is functioning well. Her intention is to turn right, and, as it happens, she turns the steering wheel and guides the car right. We are assuming here that had Sally formed an intention not to go right but, for example, to drive left instead, she would have been able to act on this intention and turn left. Thus, we are assuming that Sally has control over the car’s movements in a way that she guides the car. The car’s movements are dependent on Sally’s actions. We can therefore state that Sally has regulative control over the car. To make the distinction between regulative and guidance control, look at the following alteration to the example. Sally drives her car as before but this time a “driver instructor” is installed – a feature that allows an instructor to steer the car should the instructor think that it is necessary. Now, Sally forms an intention of turning right and acts on it, so everything seems to be as in the original example. But, if Sally had showed a sign of turning left, the driver instructor would have taken the control over the car and steered it right. As Sally actually steers her car right (without any intervention), we are inclined to say that she freely guides her car even though she is not able to drive the car into any other direction because of the presence of the driver instructor. Guidance control does not thus entail that the agent be able to act otherwise. Rather, an agent exhibits guidance control when she acts freely. Therefore, it seems that the freedom to do otherwise is not the freedom pertinent to free will because we are convinced that Sally freely steers her car right despite she could have not done otherwise. Fischer and Ravizza conclude that Sally has guidance control over her car in the latter case, whereas she exercises regulative control in the former case. (Ibid., 30—2.)

To make their case, Fischer and Ravizza harness Frankfurt examples to their assistance. In his article “Alternate Possibilities and Moral Responsibility” (1969), Harry Frankfurt described an example showing that the freedom to do otherwise is not necessary to accommodate the requirements of moral responsibility. This example is similar to the case introduced above, but here we are inclined to make a moral judgment about agents while Sally’s action can be thought to lack a moral dimension. To illustrate: suppose that there is a malicious friend of Jamie, called Carl, who wants Jamie to shoot Paul. Carl is well aware of the fact that Jamie has meditated for a long while to shoot Paul and finally concluded to do it. Carl has an ability to tinker Jamie's mind so that should

he refrain from doing it at the crucial point, Carl will use his abilities and have Jamie shot Paul. Nevertheless, Carl does not want to reveal these skills in vain, and therefore he will use these skills only if he is under the impression that Jamie will get cold feet. Carl is excellent in reading Jamie's movements of mind. Jamie shoots Paul. No intervention was needed. (Ibid., 29—30.) We can quite confidently contend, again, that the agent in question (Jamie) is in control of his actions despite the fact that he cannot but do the action at issue. Thus, an agent's acting freely (i.e. guidance control) is what grounds her moral responsibility and not the ability to do otherwise (i.e. regulative control). As Fischer and Ravizza put it: "Whereas we may intuitively suppose that regulative control always comes with guidance control, it is not regulative control, at the deep level, that grounds moral responsibility." (Ibid., 33.) In this passage, Fischer and Ravizza point out a noteworthy observation that regulative and guidance control usually come together, but in Frankfurt examples, the agent exhibits solely guidance control. Even though we do not meet cases like this in our everyday life, the power of the implications derived from them is not less.

If we grant the success of Frankfurt examples in showing that guidance control is the control pertinent to moral responsibility, we are in a position to advance to the main conclusion that can be drawn from Frankfurt examples. Recall the first premise of the indirect argument for incompatibilism: if one acts of one's own will, one must have power over one's actions. That is, if one is powerless over doing otherwise, it seems that this is not a problem for us, for it is not the ability to do otherwise that renders us morally responsible. So, if we hold that the definition of freedom of will is the ability to do otherwise, we can reject the incompatibilist argument by arguing that the ability to do otherwise is not necessary for moral responsibility. (Ibid., 34.) Further, Fischer and Ravizza maintain that the ability to do otherwise is not necessary for free agency and moral responsibility, which calls into question their compatibilism, for is it not the compatibility of free will (understood as the freedom to do otherwise) and determinism that is the main thesis of compatibilism? Fischer and Ravizza argue for semicompatibilism, holding that determinism is compatible with moral responsibility even if determinism is not compatible with the freedom to do otherwise. (Ibid., 52—3.)

A large advancement is gained from Frankfurt examples, so it might be in order to look at the implications in more detail. I will not pursue a detailed scrutiny of the success of Frankfurt examples, but instead I bring out some thoughts of my own. Nothing conflicts with the view that Frankfurt examples are consistent with determinism which would in turn imply that guidance control is consistent with determinism. (Ibid., 34.) In my opinion, this seems plausible. However, my main worry concerns an incompatibilist's possibility to argue that in our examples both Sally and Jamie are causally determined to do the actions in question because causal determinism rules out an agent's capability to not only do otherwise but also to be the source of her action. That is, I

believe the incompatibilist could argue that, for instance, Sally steers the car right in every possible relevant world (i.e. the past and the laws of nature causes her to do so) where causal determinism obtains regardless of the existence of the instructor. Further, she could claim that the possible intervener is only confusing our thinking here. “How does the possible intervener alter the situation if the agent is being causally determined to do the action at issue?” could the incompatibilist ask. Even though it can fairly be judged that the agents act freely in Frankfurt examples, the incompatibilist could equally judge that the agents are no exceptions of the ordinary cases of causally determined agents. She would thereby consent to a different kind of argument than the indirect argument that is put forth in the introductory chapter. The incompatibilist might run the following kind of argument. (1) In order for one to be morally responsible for one’s actions, one must be the ultimate source of one’s actions. (2) If determinism obtains, no one is the ultimate source of her actions. (3) Thus, if determinism obtains, no one is morally responsible for her actions. (McKenna, 2009.) I believe that this, the source argument for incompatibilism, is what the incompatibilist is pursuing when she points out that every individual is causally determined to do the actions she does. She would argue for a view stating that determinism may not rule out moral responsibility by way of ruling out regulative control. Rather, determinism is altogether incompatible with moral responsibility.

However, we are operating with the indirect argument for incompatibilism and for that reason we will leave the source argument for incompatibilism aside. Yet, it is worthwhile to keep in mind that the incompatibilist has a different sort of argument in support of her viewpoint. To go back to the indirect challenge to compatibilism, we have successfully addressed the problem that determinism poses. Since we have argued for the possibility of moral responsibility, it is our burden to provide a positive account of moral responsibility. For the present purposes, it goes by the name of guidance control.

2.2. The Methodology

Fischer and Ravizza begin their job by defining guidance control as reasons-responsiveness. It may first help to think about cases in which agents are not responsive to reasons. There are agents who act in a fixed way no matter what reasons were brought up to them. These cases include, for example, hypnotism, brainwashing and various sorts of manias and phobias. Consider Archie who has gone through hypnosis in order to kick the habit of smoking. Instead of inducing a smoke-free future for Archie, the hypnotherapist has induced an irresistible urge to steal his mate’s wallet. As things unfold, Archie steals his best friend’s wallet. However, Archie is not morally responsible because of the irresistible urge induced by means of hypnosis. Further, it is the case that no matter what reasons were brought up to him, he cannot but steal the wallet as a result of the irresistible

urge. Archie is not, as Fischer and Ravizza put it, “sensitive to reasons in the appropriate way.” (Fischer & Ravizza 1998, 36.) Cases of hypnosis, brainwashing and the like differ from those involving an unimpaired deliberation in that if sufficient reasons were presented to act otherwise, an agent acting in “normal” circumstances would therefore act otherwise. Archie and the likes can thus be distinguished from stable deliberators by reasons-responsiveness. That is, when we hold people morally responsible, we usually think that they are at some level responsive to reasons.

Although there are simple cases of impaired reasons-responsiveness as Archie, we have witnessed cases where the agent acts freely but may not be responsive to reasons. Recall Jamie who assassinates hapless Paul. It applies to Jamie as well that no matter what reasons were brought up to him, he cannot but do the deed. Jamie is not responsive to reasons, but he is, according to us, responsible for the act as we contended earlier. It seems that if an agent is to be responsive to reasons, she needs to exhibit regulative control that is ruled out in Frankfurt examples. How do Fischer and Ravizza explain moral responsibility in Frankfurt examples? First, they suggest an actual-sequence analysis that concentrates on the actual course of events whereby alternative possibilities are not required for moral responsibility. That is, when we consider more carefully the way things come about in the actual sequence of events, we can neglect the alternative scenario that results in the agent’s inability to do otherwise.

The second problem in Frankfurt examples for Fischer and Ravizza is to explain the observation that the agent is not responsive to reasons in the actual sequence of events. For that task, Fischer and Ravizza develop a mechanism-based analysis of guidance control according to which we should switch the focus from the agents to the mechanisms that actually lead to the action. Mechanisms are the processes that lead to an action, and since we have seen that an agent’s regulative control and reasons-responsiveness can be undermined regardless of the agent (e.g. interveners in Frankfurt examples), moral responsibility needs to be analyzed by the pertinent mechanisms rather than by the agents. Thus, even though an agent is not reasons-responsive, the mechanism issuing in the agent’s action may be reasons-responsive. (Ibid., 37—8.)

Let us consider once again our Frankfurt example. Jamie deliberates coolly, and decides to carry out his plan to shoot Paul. In the actual scenario, we are inclined to think that the mechanism that is operating is responsive to reasons. That is, had there been a sufficient reason to do otherwise, the mechanism operating in the actual sequence would have produced (or tried to produce) a different kind of action. In the alternative scenario, a different mechanism, Carl stimulating Jamie’s brain, would have operated, which is not responsive to reasons. So even though Jamie is not responsive to reasons in the actual sequence, his action issues from a mechanism that is responsive to reasons and thus exhibits guidance control. (Ibid.)

A few general observations should be made about mechanism-based actual-sequence reasons-

responsiveness before we go on to look at the details of this analysis. First, by developing an actual-sequence approach Fischer and Ravizza wish to emphasize that in assessing moral responsibility we should concentrate on the actual series of events. It may help to think about how, say, Jamie's mechanism operates in the actual sequence in the worlds relevant in assessing reasons-responsiveness of his mechanism if the counterfactual interveners are "filtered out." In these worlds, the mechanism of the actual sequence would result in a different action were there sufficient reasons brought up. (Ibid., 38.)

Second, by taking a mechanism-based strategy to describe guidance control, Fischer and Ravizza invite us to focus on the pertinent process that leads to an action. As they put it: "Note that, although we employ the term 'mechanism,' we do *not* mean to point to anything over and above the process that leads to the relevant upshot; instead of talking about mechanism that leads to (say) an action, we could instead talk about the process that leads to the action, or the 'way the action comes about.'" (Ibid.) It is the relevant history of how things come about, then, that is pointed out by referring to a mechanism producing an action. Further, the mechanism-based strategy differs from the agent-based theories, which focus solely on the agent. The agent-based theories contain a counter-intuitive result: given that Jamie acts freely, the agent-based theories find, however, his moral responsibility undermined because he is not able to act otherwise. That is, we can picture a scenario (or a possible world) that is the same into the smallest detail as the one in our Frankfurt example but the scenario does not contain Carl. In this scenario, should Jamie decide to refrain from murdering Paul, he would be able to act upon his decision. The agent-based theories do not test the agent's pertinent psychological capacity through the possible worlds, so to speak. Therefore, what we attain by concentrating on the mechanism operating is that we do not need to require the agent to possess regulative control. Thus, we can evade the troublesome indirect argument for incompatibilism since its problematic nature arises from the claim that no one can do otherwise.

To sum up, Fischer and Ravizza are advising an analysis of moral responsibility that focuses on the actual history of the events that result in an action. Looking to the actual events of Frankfurt examples, we notice that moral responsibility is not dependent on the ability to act otherwise since the agent's action issues from a mechanism that exercises guidance control although he could not have done otherwise. By advancing a mechanism-based analysis, we see that although *the agent* is not responsive to reasons, her *mechanism* can be exercising guidance control rendering her responsive to reasons and morally responsible.

2.3. Moderate Reasons-Responsiveness

Having learned the characteristics of the reasons-responsiveness theory of moral responsibility, we can begin to study it more thoroughly. Fischer and Ravizza aim to describe an account of moral

responsibility that fulfills the two following conditions. First, they provide a theory that grounds an agent's moral responsibility in her reasons for action. Second, as an implication of the first condition, they are able to offer a theory that explains how one is morally responsible for doing knowingly wrong. In order to find the right degree of reasons-responsiveness sufficient and necessary for moral responsibility, Fischer and Ravizza introduce two extremes: strong and weak reasons-responsiveness (hereinafter SRR and WRR). We shall begin with SRR. "SRR obtains under the following conditions: if K [certain kind K of mechanism] were to operate and there were sufficient reason to do otherwise, the agent would *recognize* the sufficient reason to do otherwise and thus *choose* to do otherwise and *do* otherwise." (Ibid., 41.) SRR can be thus analyzed in three separate conditions. The first condition of SRR involves a sufficient receptivity to reasons. That is, the agent that operates from an SRR mechanism needs to recognize the reasons there are and whether those are sufficient reasons for acting otherwise. The second condition involves reactivity to reasons. In order for an agent to act from an SRR mechanism, she needs to choose accordingly with the sufficient reasons. The third condition consists in translating one's choice into action. So, to act from an SRR mechanism one must finally, after recognizing the sufficient reasons and choosing according with those reasons, act according with her choice. (Ibid., 41—2.)

Although this sort of tight fit between reasons and actions would be desirable for some behavior, it is not necessary for moral responsibility. To see why, consider the following example of a weak-willed agent. After a normal unimpaired reasoning Jones elects to go to a football match for which he has a sufficient reason, and as things happen in the actual sequence, he goes to the game. Suppose that if there had been a sufficient reason not to go, Jones would have been weak-willed. Further, if Jones had had a sufficient reason not to go to the game, it would have been that his girlfriend, Mary, is ill. So, had Mary been under the weather, Jones would have decided to go to the game, even though he would have recognized the sufficient reason not to go to the game. Now that we study Jones's behavior, we see that he is not strongly responsive to reasons. He would recognize the reason for staying with Mary, but he would not meet the second and third condition. That is, he would not choose according with the sufficient reason to stay with Mary and, in consequence, he would not act according with the choice. Therefore, if we accept SRR as necessary for moral responsibility, we would deem Jones to lack moral responsibility. This is an odd implication because, clearly, moral responsibility should cover blameworthy actions that stem from weakness of the will. (Ibid., 42—3.) SRR cannot thus describe Fischer and Ravizza's second aim that is explaining how one is morally responsible for doing knowingly wrong.

The troubles that arise with SRR give reason for a looser fit between reasons and actions. WRR is the opposite extreme that works under looser requirements: "In contrast, under weak reasons-responsiveness, we (again) hold fixed the actual kind of mechanism, and we then simply require

that there exist *some* possible scenario (or possible world) in which there is a sufficient reason to do otherwise, the agent recognizes this reason, and the agent does otherwise.” (Ibid., 44.) Let us consider the example of Jones again. There are certainly some possible scenarios in which Jones would not go to the game. For example, if Mary would have selflessly stretched for Jones's favorite board game on the shelf so that his tending to Mary would not be that dull, and as a result broken her leg. Jones’s responding to some possible circumstances, which surpass his disposition toward weakness of the will, proves that he acts from a WRR mechanism. (Ibid., 45.)

Even though WRR serves as the basis for guidance control better than SRR, it still allows bizarre patterns of acting to satisfy the control condition for moral responsibility. Consider a maniac who slaughters the next neighbor he sees in every scenario but one. In this one scenario, he refrains from this cruelty if the neighbor sings the song “For He's a Jolly Good Fellow” to the maniac. The maniac goes out to the corridor, and as the next unsuspecting neighbor comes, he murders her since she did not sing the song. The maniac of our story acts from a WRR mechanism, for were the neighbor to sing the song, he would have responded to the reason not to kill her and acted accordingly. However, it seems that the maniac has a very vague type of control over his actions. WRR is thus too weak to describe the control pertinent for moral responsibility. (Ibid., 65—6.) Further, the first condition (presented in the beginning of this section) that Fischer and Ravizza aim to fulfill is not satisfied since one acting from a WRR mechanism may not have an adequate connection between her reasons and actions.

We need to refine our reasons-responsiveness theory so that we will avoid the pitfalls that SRR and WRR involve. Moderate reasons-responsiveness (hereinafter MRR) is an improvement which avoids the all too stringent conditions of SRR and tightens up the loose conditions of WRR. Fischer and Ravizza analyze MRR by distinguishing two stages of reasons-responsiveness, which are introduced along with SRR. The first condition of SRR is the sufficient receptivity to reasons. That is, recognizing the reasons there are. Contrary to SRR, what is required for one to act from an MRR mechanism is not to recognize all the sufficient reasons to do otherwise, but instead the mechanism that operates must show a pattern of rationality in recognizing the reasons. (Ibid., 69—70.) Recall the example of Jones and her girlfriend, Mary. Suppose that, as before, if Mary had broken her leg, Jones would not have gone to the game. But consider that Jones’s receptivity to reasons is straitened so that he would not have recognized as a sufficient reason not to go to the game had Mary broken her neck, suffered a concussion or got her shoulder dislocated. Suppose further that, Jones would only have recognized the reason for not going to the game if Mary had broken her leg. Jones’s behavior seems thus similar to the maniac who lurks in the corridor. Therefore, Fischer and Ravizza maintain that “in order for an agent to be morally responsible for an action, then, the actual mechanism that issues in his action must be at least ‘regularly’ receptive to reasons.” (Ibid., 71.)

The regularity involves that an agent recognizes how reasons hang together. So, if Jones would have recognized Mary's breaking her leg as a reason to act otherwise, he must have recognized her breaking her neck, suffering a concussion or getting her shoulder dislocated as a reason to act otherwise as well for him to act from an appropriately reasons-receptive mechanism.

Recall the second and third conditions that constitute a SRR mechanism. That is, choosing and acting in accord with the strongly receptive mechanism. Properly modified, they make up the second stage of moderate reasons-responsiveness, the reactivity component. Whereas regular receptivity to reasons is sufficient for an MRR mechanism, one needs to be only weakly reactive to reasons in order to act from a moderately reasons-responsive mechanism. Weak reasons-reactivity, Fischer and Ravizza philosophize, consists in an agent's mechanism's reacting to *some* incentive to act otherwise: "Our contention, then, is that a mechanism's reacting differently to a sufficient reason to do otherwise in some other possible world shows that the same kind of mechanism can react differently to the *actual* reason to do otherwise." (Ibid., 73.) Consider Jones once again. For him to act from a regularly reasons-receptive mechanism, he needs to *recognize* as a sufficient reason for acting otherwise that Mary had suffered a concussion, got her shoulder dislocated, broken her leg or neck, and the like. However, in order for him to act from an MRR mechanism, he actually needs to *react* to only one of these reasons, such as Mary's breaking her leg. As a consequence, we see that Fischer and Ravizza argue for a stark asymmetry, which involves that one must not be as equally reactive as receptive to reasons to act from an MRR mechanism. They motivate their asymmetry thesis by claiming that "[one] must simply display *some* reactivity, in order to render it plausible that his mechanism has the 'executive power' to react to the actual incentive to do otherwise." (Ibid., 75.)

In my opinion, the asymmetry thesis is problematic. To see why, it may help to think about the protagonist of our example, Jones. Suppose that he would recognize the reasons presented above as sufficient for acting otherwise. But, since he would react only to one reason, Mary's breaking her leg, it is hard to see how this would serve as a basis for his moral responsibility. There must be some other similar reasons to which an agent acting from an MRR mechanism reacts. Otherwise, agents with bizarre pattern of reactivity will qualify as morally responsible agents. Why do Fischer and Ravizza adopt such an asymmetry? I believe the reason for this is that it allows them to explain agents that knowingly do wrong, for they require regular receptivity but only weak reactivity to reasons. That is to say, morally blameworthy action consists in, first, recognizing the reasons to act otherwise and, secondly, failing to act in accord with those reasons. Therefore, say, Jones is receptive to a range of reasons that ought to move him to stay with Mary and, thus, he knows that he has a sufficient reason to stay with Mary. However, since he does not react to all of these reasons, he goes to the game. Would Mary have had her neck broken and Jones would have gone to

the game (because he reacts only to the reason that Mary would have had her leg broken), it seems that Jones is doing knowingly morally wrong and is subject to blame. Yet, Jones's being morally responsible is unclear; if he reacts only to one reason to act otherwise, his behavior is suggesting that he may not be subject to blame. That is, his peculiar behavior does not intuitively meet the standards that one would assign to a morally responsible agent, which, in my opinion, gives grounds for doubting his general capacity to exercise guidance control.

What should Fischer and Ravizza do to remove this unwanted consequence? In my opinion, they should explain the range of cases in which an agent recognizes but does not react to reasons to act otherwise. What this means is that they should describe all the cases in which an agent fails to react to an incentive to act otherwise but could still be counted responsible. For instance, in Jones's case, those who endorse the asymmetry thesis could better their positions by explicating the instances and reasons in which Jones ought to act otherwise but fails to do that – Jones does not refrain from going to the game even if Mary breaks her neck because Jones is weak-willed and selfish, etc. Therefore, he is subject to blame not because he stays with Mary only in one another scenario in the actual world but because, say, his action-producing mechanism is corrupt. Why would this do the trick? By weak reactivity, Fischer and Ravizza are pursuing a principled description that connects a person's reasons to her actions. However, it is too general. It allows people with idiosyncratic patterns of reactivity to be treated as morally responsible. Instead, were they to name all the instances in which an agent does not react to a reason he recognizes as sufficient for acting otherwise, they would successfully connect an agent's reasons to her actions.

One could ponder why the pattern of reactivity could not be similar to the receptivity component, that is, regularly reactive? It must be borne in mind that some gap must be left between the reasons an agent recognizes and reacts to. For if one were as receptive as reactive to reasons, it would entail that there would be no room for blameworthy action since this sort of description of moral responsibility would require that an individual reacts to all regularly recognized reasons. The problem is the same that we encountered when sketching SRR. So if Fischer and Ravizza could explain the range of cases where an agent recognizes the reason to act otherwise but does not react to those, it would still allow them to explain morally blameworthy action, but at the same time, they could steer clear of the current problematic asymmetry thesis.

I have not yet said anything about MRR with respect to incompatibilism. The compatibilist's strategy goes on to describe how the action individuation takes place. In Fischer and Ravizza's study, it is explained by way of expressing certain reaction to the reason in some other relevant possible world, which proves that the agent has the cognitive capacity pertinent to moral responsibility. Transparently, the incompatibilist will not be persuaded that this is the way of progress. "How is it fair to hold an agent morally responsible if she acts differently in some other

possible world were some other reasons brought up?” asks the incompatibilist. The incompatibilist will demand that the reason that moves the agent to act differently in some other possible world should move her to act differently in the actual world. However, the actual world is not, the incompatibilist continues, such that the reason present in a relevant possible world would serve as a basis for acting differently in the actual world. Therefore, the incompatibilist asserts that the agent must exhibit regulative control in order for her to be morally responsible for her behavior since she must be able to act according to the reason that moves her to act differently in some other possible world. That is, she must be able to act otherwise.

2.4. Taking Responsibility

Despite the problems that arise along with formulating an adequate analysis of guidance control, we can grant that Fischer and Ravizza successfully offer an insightful scrutiny of guidance control as MRR, and therefore we can move on to the next component of their theory of moral responsibility: an MRR mechanism needs to be an agent's own. The following kind of distinction is useful here. The nature of moral responsibility divides the compatibilist field into two separate camps. Those arguing for a nonhistorical model of moral responsibility maintain that an agent's moral responsibility is judged by her “snapshot” properties. That is, an agent's moral responsibility can be traced to her actual psychological properties. (Ibid., 184.) In contrast to the nonhistorical approach, a historical compatibilist contends that moral responsibility is a historical phenomenon. According to historical compatibilists, moral responsibility cannot be evaluated only by an agent's “current time-slice properties,” but it is necessary to take into account the historical process through which the agent takes responsibility. (Ibid., 196.)

Recall the distinction between the agent- and the mechanism-based approaches. The distinction is useful here as well, since those arguing for the mechanism-based approach reject the nonhistorical view. The mechanism-based approach maintain that in order for one to be morally responsible, one must act from an MRR mechanism which renders the agent morally responsible because one acts on guidance control. Guidance control is based upon an agent's psychological capacity rather than her access to alternative possibilities. As we study her psychological capacity, we look at “the way the action comes about” rather than her current time-slice properties (McKenna, 2009). As ought to be apparent, Fischer and Ravizza endorse the historical approach to moral responsibility.

The following example will give further support to the historical approach to moral responsibility. Suppose that Jones goes to the game and this action is produced by an MRR mechanism. Furthermore, a group of experimental neurologists wanted to pilot their new brain enhancer program installing an artificial psychological structure in the agent. By accident, Jones

was elected for this duty and he received an MRR mechanism in an overnight process. Suppose that Jones was in no control whether to be chosen to this pilot program, and the next morning Jones felt just like every other morning; that is, the experiment was a success as its main goal is to leave the agent “untouched.” Can we now plausibly say that Jones is acting on guidance control as he goes to the game? It would seem that the mechanism issuing in an action is not Jones's, and the group of neurologists should instead be held responsible for the actions produced by the MRR mechanism installed in Jones. The historical compatibilists contend specifically that as the mechanism issuing in going to the game is not Jones's, he cannot be held responsible, for it is a crucial ingredient of moral responsibility that the mechanism issuing in an action be an agent's own. However, the nonhistorical compatibilists endorse the idea that the way an agent comes to have the specific psychological constitution does not matter when judging her moral responsibility. Thus, Jones is deemed morally responsible by the nonhistorical compatibilists. And that is a startling conclusion. (Fischer & Ravizza 1998, 197–201.)

This raises the question of what leads a nonhistorical compatibilist to make a concession of this size. The answer is as follows: if a compatibilist holds that the history of an agent needs to be certain in order for her to be morally responsible, then the compatibilist allows that there are histories that undermine an agent's moral responsibility. According to the incompatibilist, it would follow from this reasoning that a deterministic history is undermining an agent's responsibility. For if manipulated agents are not morally responsible by virtue of their past, a deterministic past would similarly undermine an agent's moral responsibility. A manipulated agent's history, the incompatibilist continues, does not differ from a normal one in a deterministic world. (McKenna, 2009.)

It makes now more sense why a compatibilist might argue for the nonhistorical approach, but I still believe that the brainwashing example sketched above offers a *prima facie* justification for the historical approach. Despite the incompatibilist's possibility to argue in the fashion set above, we can develop a historical notion that shows how an agent takes responsibility. That is, we must present the pertinent conditions that contribute to an agent's taking an MRR mechanism to be her own. We will resist the contention that a deterministic history would be similar to a manipulated agent's history. By way of showing that an agent takes an MRR mechanism to be her own, it can be argued that incompatibilists confuse deterministic causation with compulsion and the like. According to a compatibilist, incompatibilists mistakenly hold that determinism entails that every individual does compulsively whatever she does. And because nothing in our treatment contradicts the obtaining of determinism, we can show that manipulated agents crucially differ from agents with a normal deterministic history.

We may now turn to Fischer and Ravizza's historical notion in more detail. They offer a three-

stage analysis whereby an agent comes to take responsibility. First, an agent must see herself as an agent, which means she needs to recognize that her actions have results in the world. “The agent thus sees that his motivational states are the causal source – in certain characteristic ways – of upshots in the world.” (Fischer & Ravizza 1998, 210—1.) The second condition involves that the agent sees herself as an apt target of the reactive attitudes, which are understood as moral expressions, say, praise and blame. That is, the agent must see it justified that others assign the reactive attitudes towards her. As Fischer and Ravizza clarify: “the second condition states that the individual must see that in certain contexts it is ‘fair,’ in the sense of being part of our given social practices, for others to subject him to the reactive attitudes in certain circumstances.” (Ibid., 211.) Note that, the certain circumstances and contexts Fischer and Ravizza refer to are understood as circumstances and contexts where the agent’s behavior gives reason to reactive attitudes, such as praise and blame. Third and finally, the agent’s view of himself, specified in the first two conditions, needs to be based on the evidence. Fischer and Ravizza develop a metaphor of moral responsibility as a sort of moral conversation stating that “our practices of holding one another responsible may be viewed (in part) as a kind of moral address.” (Ibid., 213—4.) Motivating the third condition, they continue: “when a person takes responsibility, he is ‘asking’ the moral community to recognize him as a legitimate participant in this moral conversation. When – but only when – this ‘request’ arises in response to his evidence [...], it indicates that the person is indeed a suitable participant in the conversation he is asking to join.” (Ibid., 214.) Thus, Fischer and Ravizza take a subjectivist approach to moral responsibility. One must see oneself as morally responsible in order to be morally responsible.

A notable consequence of Fischer and Ravizza’s subjectivist notion concerns the second condition of taking responsibility. They distinguish between nonreflective and reflective individuals. Most individuals belong to the former group, and they are required to form a judgment about the fairness of our social practices as described above. As for the latter group, the social judgment is not sufficient. The reflective individuals, such as some philosophers ruminating on the relationship between determinism and the fairness of applying the reactive attitudes, need to form “some sort of metaphysical judgment to the effect that, all things considered, it is appropriate to subject certain individuals – including himself – to the reactive attitudes in certain contexts.” (Ibid., 226.) The reflective individuals may thus opt out of the social practices of assigning the reactive attitudes by pointing out that determinism would rule out the meaningfulness of these practices. Fischer and Ravizza accept that this group is not responsible for their actions since they fail to meet the second condition of taking responsibility, given that determinism rules out moral responsibility (ibid., 228).

This may be too a permissive stance, and I would suggest that these individuals should extend

their skeptic view to their everyday social life. I am quite certain that the price of being excluded from the moral practices is too high even for the most stubborn hard determinists. Their communities would begin to treat them as dangerous animals and the skeptics would live a miserable life without any moral relationships. So, I suggest that if the reflective individuals opt out of moral practices because no legitimate metaphysical judgment is brought up to them, they should run empirical tests on their hypothesis. Even though Fischer and Ravizza do not demand skeptics to test their view in everyday life directly, they note that “there are strong incentives not to opt out of moral responsibility. Agents who genuinely fail to take responsibility – and thus view themselves as lacking control – are legitimately sequestered from society, and are deprived of the opportunity to participate in the moral community.” (Ibid., 229.)

Having presented guidance control in its entirety, we may view the brainwashing case in the light of the conditions set above. The third condition of an MRR mechanism’s being an agent’s own is designed to block the possibility of an agent being electronically manipulated into taking responsibility (ibid., 235—6). That is, Jones’s MRR mechanism is not his own because his seeing himself as an agent and apt target of the reactive attitudes is not based upon his evidence for these beliefs. Therefore, despite Jones’s mechanism issuing in the going to the game is moderately responsive to reasons, he has not taken responsibility for the mechanism. For that reason, it is not plausible to hold him morally responsible for neglecting his girlfriend. However, that the first two conditions need to be grounded in the evidence may not be sufficient to exclude yet more refined examples of agents manipulated into taking responsibility. Fischer and Ravizza are aware of this and therefore point out that “the agent’s view of himself must be based on his evidence *in an appropriate way* [emphasis added].” (Ibid., 236.) As we will see in the last section, the notion of appropriateness is contested, and it might eventually be up to the reader to decide whether the manipulated agent (or any of us) views herself as morally responsible in an appropriate way. I will now turn to Mele’s compatibilist proposal where further comments on Fischer and Ravizza’s analysis of guidance control as reasons-responsiveness come up.

3. History-Sensitive Compatibilism

In studying Alfred Mele’s compatibilist notion, I will proceed in two stages. First, I will bring up his critique to Fischer and Ravizza entailing that moral responsibility should not be analyzed in terms of the reactive attitudes. Even though he himself remains agnostic about compatibilism, our second task is to describe his compatibilist proposal.

3.1. A Critique of Fischer and Ravizza

Mele claims that the necessary condition for moral responsibility that Fischer and Ravizza advance

is too strong and their sufficient conditions are too weak (Mele, 2006 146). I will first take up the critique to the necessary condition, which is familiar to us from the preceding section: an agent's mechanism issuing in an action must be her own. Recall the second condition for taking responsibility: the agent must see it as fair for others to subject him to the reactive attitudes. Their subjectivist account thus leads to the possibility that certain reflective individuals, such as hard determinists, jettison the fairness of being a subject to the reflective attitudes. In other words, the reflective individual denies her being morally responsible. It is worth underscoring that Fischer and Ravizza admit that this peculiar party is out of the boundaries of moral responsibility, and thereby they believe that the gain for opting out of moral responsibility is notably less than being a part of the moral practices.

I concluded that these individuals should test their hypothesis of the metaphysical nature of the world in light of the consequences it entails. Mele develops this reasoning a little further. He invites us to consider the following example of a philosopher, Phil. Phil is an ex-compatibilist converted to hard determinism, but the conversion to hard determinism does not change his life drastically. The conversion did not include any sort of manipulation or processes that are out of his control. Moreover, he is merely intellectually committed to determinism and continues to value others' feelings and believe that relationships are big part of his life. By hypothesis, Phil is mistaken about the nature of the world, and morally responsible agents are abundant in the world. As any of us, he occasionally tells small lies to make his life easier. Nevertheless, Phil is benign and, in general, treats people well. As far as his behavior, nothing has changed since his conversion to hard determinism. (Ibid., 147.)

What Mele wishes to emphasize with the example is that the second condition of taking responsibility is too strong. That is, even though Phil does not satisfy the second condition, it would seem implausible not to hold him morally responsible for the praiseworthy and blameworthy actions at question. (Ibid., 147—8.) So whereas I was suggesting that Fischer and Ravizza should tighten up their second condition by adding the application to everyday life clause, Mele contends that they should loosen the condition. The difference is this: I believe that their theory of taking responsibility, when it comes to hard determinists, is not working in a real life, while Mele thinks that the theoretical outcome is more interesting than the social outcome. I think that if hard determinists apply their theoretical beliefs into their social lives, they run into troubles. But, few of them would extend their theoretical beliefs concerning their social relationships. As Phil, after becoming convinced of incompatibilism's superiority and finding determinism true, they would hardly begin to act immorally, owing to the discovery. The point Mele wants to make is that moral responsibility should not be analyzed in terms of the reactive attitudes. As for this argument, it is Fischer and Ravizza's subjective thesis of moral responsibility that he criticizes.

According to Mele, another problem of MRR is that the sufficient condition for it is too weak. His concern is weak reasons-reactivity. (Ibid., 149.) Recall that reactivity component consists in the capacity to translate reasons into choices, and the weak version Fischer and Ravizza advance entails that although an agent's mechanism does not react to a reason to act otherwise in the actual world, there is some possible world in which the agent's mechanism operates and acts otherwise for that reason. Mele motivates his assertion that weak reasons-reactivity is too weak by presenting the following example.

Fred's agoraphobia is so powerful that he has not ventured out of his house in ten years, despite his family's many attempts to persuade him to do so and the many incentives they have offered him. Owing to his fear, he often has decided not to do things that he believed he had good reason to do (and then behaved accordingly), including some things that he believed he morally ought to do. For example, he recently decided to stay at home (and stayed there) rather than attend his beloved daughter's wedding in the church next door. Now, in some possible world with the same laws, there was a raging fire in Fred's house on his daughter's wedding day. Fred, it turns out, is even more afraid of raging fires than of leaving his house. Judging that he had a good reason to leave his house, he decided to do so, for that reason; and then, making a heroic effort, he walked next door to the church. (Ibid., 150.)

Nothing suggests that Fred's mechanism issuing in the action in the actual world could not be regularly receptive to reasons and be his own, so it can be supposed to have these two features. Therefore, on Fischer and Ravizza's account, Fred is deemed morally responsible. However, Mele argues for a more intuitive contention that Fred falls short of moral responsibility due to his severe phobia. (Ibid., 150.)

Mele wants to point out that even in extreme cases of phobia or the like there is usually some reason that the agent would recognize and do otherwise for the reason. According to Mele, this indicates that Fischer and Ravizza should tighten the weak reasons-reactivity condition in order to exclude Fred from being morally responsible for missing the wedding. (Ibid., 152.) The problem I observed with the weak reasons-reactivity is that it would seem bizarre if an agent reacts only in one possible world to the reason to act otherwise and would thus act otherwise in the one possible world rendering her morally responsible in the actual world. Mele notices the same problem and illustrates it strikingly with the example of Fred. On the one hand, I suggested that Fischer and Ravizza should explain the sphere of reasons an agent recognizes but does not react to in order for them to explain those who knowingly do wrong. On the other hand, Mele contends that weak reactivity to reasons is too weak since agents like Fred are plausibly held morally responsible. Both of us agree that some tightening would be in order. Instead of explaining blameworthy action, Mele wants to see that the reactivity component would exclude individuals with severe phobias or addictions from being morally responsible for some actions. Where is the difference? I believe that

explaining blameworthy action and compulsive behavior are two sides of the same coin in this matter. That is, in my opinion, the weak reasons-reactivity cannot do the job of connecting an agent's reasons for action (and for not acting, as in Fred's case) properly as becomes apparent with the example of Fred and an agent that reacts to but one reason to do otherwise. However, to find a principled way to remove these peculiarities is not an easy job.

Finally, I will bring out, in short, some defenses Fischer and Ravizza might pursue in reply to these critiques. They note that being morally responsible does not necessary entail being praiseworthy or blameworthy. (Ibid., 150.) Thus, Fred would be morally responsible for missing the wedding but not blameworthy for that. This would remove the conclusion that conflicts our intuitions about Fred. Nevertheless, it does not change the outcome that, according to Mele, Phil is more clearly subject to both praise for some of his good actions and blame for some of his bad deeds than Fred is. (Ibid., 152.) Yet, on Fischer and Ravizza's account, Phil is not morally responsible. How could Fischer and Ravizza motivate that? Mele speculates with the possibility that Phil might hold himself an inappropriate target to the reactive attitudes because he believes that the mechanism issuing in an action would not be even weakly reactive to reasons. Therefore, contrary to failing to meet the "taking responsibility" condition, Phil would fail to act from a weakly reasons-responsive mechanism and be for that reason lacking moral responsibility. However, were he to opt out of moral responsibility on the grounds that he is convinced either that he cannot recognize the reasons to act otherwise or that he is capable of recognizing some of them but not acting on them, "his action-producing mechanisms are frighteningly inflexible and limiting." (Ibid., 153.) For instance, Phil would perhaps recognize as a sufficient reason to act otherwise should his neighbor be about to die from a cardiac arrest in five meters away in a possible world with the same laws of nature. Nevertheless, he would not react to this reason in the actual world because he would be convinced that either he does not recognize the reason to help the neighbor or that he recognizes the reason to act otherwise but is incapable of acting on this reason. Phil is not, however, this sort of person as he cares about pleasure and well-being of his and others. According to Mele, it is evident that the reflective individuals would not opt out of weak reasons-reactivity, which results in a fatalistic behavior, such as that above (ibid.). In summary, Mele maintains that Fischer and Ravizza cannot do the job of describing an adequate account of moral responsibility since their treatment has the consequence that severely phobic and compulsive behavior is deemed morally responsible.

3.2. Conditions for Free Agency

As I mentioned earlier, many compatibilists hold that where incompatibilists go astray, is that they conflate action in a deterministic world with compelled action. Mele holds that Fischer and Ravizza's notion of moral responsibility defined by way of possible worlds offers a strong basis for

demarcating uncompelled and compelled action in a deterministic world (ibid., 156). With the refinements to conditions on MRR adduced, Mele contends, it would be a plausible description of moral responsibility. However, because Mele does not grant the second condition on taking responsibility, he provides a compatibilist set of sufficient conditions for free agency or, in Fischer and Ravizza's terms, for an agent to take an MRR mechanism to be her own.

Mele illustrates his notion that free agency is dependent on an agent's history with the following example of a manipulated agent.

Ann is a free agent and an exceptionally industrious philosopher. She puts in twelve solid hours a day, seven days a week, and she enjoys almost every minute of it. Beth, an equally talented colleague, values many things above philosophy for reasons that she has refined and endorsed on the basis of careful critical reflection over many years. Beth identifies with and enjoys her own way of life, and she is confident that it has a breadth, depth, and richness that long days in the office would destroy. Their dean wants Beth to be like Ann. Normal modes of persuasion having failed, he decides to circumvent Beth's agency. Without the knowledge of either philosopher, he hires a team of psychologists to determine what makes Ann tick and a team of new-wave brainwashers to make Beth like Ann. The psychologists decide that Ann's peculiar hierarchy of values accounts for her productivity, and the brainwashers instill the same hierarchy in Beth while eradicating all competing values—via new-wave brainwashing, of course. Beth is now, in the relevant respect, a “psychological twin” of Ann. She is an industrious philosopher who thoroughly enjoys and highly values her philosophical work. Largely as a result of Beth's new hierarchy of values, whatever upshot Ann's critical reflection about her own values and priorities would have, the same is true of critical reflection by Beth. Her critical reflection, like Ann's, fully supports her new style of life. (Ibid., 164—5.)

The process how Beth comes to have the new values contains an external intervener, and this violates her autonomy rendering her unfree. We may consider the difference between Ann and Beth as the difference between caused and compelled behavior (Ibid., 166). So, the question is, how does Mele explain the difference between Ann and Beth? In general, how is it that the history matter?

A central plank of Mele's theory of free agency is the conception of psychological autonomy. First, it is conceivable that an agent does not possess psychological autonomy in performing an action *A* if “he expresses unsheddable values in A-ing.” (Ibid., 171.) That is, an agent may even manipulated act freely if she is able to shed her newly implanted values. Therefore, for example, Beth can retain her autonomy if she can reflect on the implanted values and elect whether to act on those values. Secondly, an agent's psychological autonomy is undermined if “owing directly to those values, he could not have done otherwise than *A* in the circumstances.” (Ibid.) Third, an agent does not exhibit psychological autonomy if “those values were very recently produced in a way that bypassed his capacities for control over his mental life by value engineering to which he did not consent and are seriously at odds with autonomously acquired values of his that were erased in the

process.” (Ibid.) The third sufficient condition on psychological autonomy requires that an agent, who is able to step back and deliberate on her values, did not acquire these by means of external forces beyond his control (ibid., 166—7). Fourth and fifth, an agent’s psychological autonomy is violated if “he retains no preexisting value that is promoted by his having the unsheddable values he expresses in *A*-ing and *A* is the first overt action he performs on the basis of his new values.” (Ibid., 171.) A conjunction of these conditions amounts to a behavior that is not free.

The set of conditions that jeopardize an agent’s autonomy is exhaustive. Note that it offers a description of the circumstances under which an agent is *not* acting freely. So it does not tell us anything specific about when an agent is free and morally responsible to do the action at issue, which is something Fischer and Ravizza pursue in their positive account of moral responsibility for actions. In order for us to grasp the details of Mele’s notion on free agency, I believe that the right strategy is to contrast the compatibilist set of sufficient conditions with the example of Ann and Beth.

Mele deals with various critiques to defend his notion that whereas Ann freely expresses her values, Beth does not. According to him, his critics overlook the two salient features of the story: Beth’s capacities for control over her mental life were bypassed and her values are unsheddable. One of his critics, Daniel Dennett, points out that it is merely Beth’s having false beliefs of her past that undermine her autonomy (ibid., 175). Dennett contends that if one, who falsely believes that she has a certain history, is not autonomous and morally responsible, then moral responsibility would depend on the truth of one’s beliefs about her past. Therefore, a praiseworthy action would begin to lack this feature by virtue of an aspect that is truly irrelevant to the action in question. Mele turns down this claim. He asserts that there are no grounds for holding that Ann’s autonomy would be undermined by her falsely believing that her values are result of, say, brainwashing. (Ibid.) Suppose that Ann takes a bank loan to finance her daughter’s college studies. Would it change the action’s praiseworthiness if she falsely believed that the values producing the action are a result of brainwashing? Mele does not believe so and points out that it is rather Beth’s not consenting to the brainwashing process that renders her unfree. (Ibid.)

The second critique Mele takes up is Nomy Arpaly’s assertion that Beth’s conversion to an industrious philosopher raises the question of how does this differ from conversions of the sort we meet in everyday life, such as an atheist finding God. Granted, to identify Beth with a more realistic conversion of values has an intuitive appeal, but Mele has an answer up his sleeve. Suppose that a party animal, who grows tired of her previous way of life, becomes an industrious worker through a process that she interprets as growing up. She reflects on her new values and meets the standard description of becoming an adult. How is this dissimilar, if at all, to Beth’s conversion to her new values that make up her new way of life? Arpaly concludes that if the conversions are similar in the

relevant respect, then, Beth is autonomous in expressing her values if Ann is (*ibid.*, 179). However, the conversions are not similar in the relevant respect. First, Beth's newly implanted values are unsheddable. As an example of an at least very modest unsheddable value, Mele offers a parent of a newborn having a value that "places killing their infant child psychologically out of bounds." (*Ibid.*, 183.) So, at least for a time after the brainwashing, Beth expresses unsheddable values whereas the agent grown tired of being a party animal does not express unsheddable values in a realistic reading of the example. Second, the party animal reflects on her new values and she is conscious of the ongoing growing up process. Her capacities for control over her mental life are not thus bypassed. This cannot be said of Beth whose newly implanted values are not a result of a critical reflection and the like. Arpaly's aim is to challenge the distinction between compelled and caused behavior. She wants to point out that if an agent's values have been produced by value engineering such as brainwashing, the agent may not be any different from someone whose moral responsibility is undermined "by some unlucky chance of a force of nature." (*Ibid.*, 179.) However, we argue that Arpaly fails on this score.

Now that we have described history-sensitive compatibilism, where does this leave us? Alfred Mele provides a set of sufficient conditions for autonomous agency and moral responsibility, not necessary and sufficient. That is, he does not give an analysis of the conditions under which an agent is autonomous but he offers constraints on it. He continues: "Motivating the assertion that conditions sufficient for free action are satisfiable by an agent in a deterministic world would turn the trick." (*Ibid.*, 164.) Therefore, in order for us to present an agent who satisfies these conditions in a deterministic world, we need to dovetail a positive account of moral responsibility, such as Fischer and Ravizza's, with Mele's constraints. An important gain in Mele's psychological autonomy is how it illuminates the difference between a deterministic history and a history constructed by brainwashers or similar indeterministic mechanisms that intervene an agent's autonomy.

4. Concluding Remarks

The examination of compatibilist-friendly moral responsibility proceeds in two steps. First, we need to address the problem determinism poses to free agency and moral responsibility after which we provide the necessary conditions on moral responsibility. In terms of reasons-responsiveness, this consists in the testing of an agent's capacities to exhibit guidance control by way of possible worlds or scenarios. An agent's mechanism issuing in an action is moderately responsive to reasons if it regularly recognizes the reasons there are and it reacts to some of these reasons. In the stage of receptivity, an agent is regularly receptive to reasons if she can exhibit a pattern of recognition that proves her ability to connect different reasons together. In our example, Jones needs to recognize as

a sufficient reason to act otherwise not only that Mary breaks her leg but also that she breaks her neck, etc. As for the stage of reactivity, an agent must only react to one of these reasons in order for her to show that she has the executive power to react to the actual incentive to do otherwise. Problems arise from the both fronts. First, if there is a sufficient reason to do otherwise that an agent recognizes and does otherwise for that reason in a possible world, no matter how remote, agents like Fred meet the receptivity condition. (Ibid., 156.) Second, Fred meets the reactivity condition as well since he is more afraid of raging fires than of going out. Therefore, he is weakly reactive to reasons because in the case of raging fire in his house he would be going to the church next door. In the actual course of events without the fire, he is thus deemed morally responsible for staying home. To address the former problem, Fischer and Ravizza should explain the relevant sphere of the possible worlds. That an agent recognizes a very remote world in which she would act otherwise does not serve as a basis for her to be sufficiently receptive to reasons. In order to answer the latter problem, Fischer and Ravizza should explain the circumstances in which an agent does not react to the reason to act otherwise even though recognizing the reason.

The second step of providing a compatibilist account on moral responsibility is to show the way in which an agent takes her mechanism issuing in an action to be her own or, in other words, how an agent exhibits psychological autonomy. The necessary conditions of moral responsibility as reasons-responsiveness are that an agent sees herself as an agent, that she is an apt target of the reactive attitudes and that the agent's view of herself as defined in the two above conditions is based on the evidence. Hence, Fischer and Ravizza argue for a subjective notion of moral responsibility: in order for one to be morally responsible, one must see oneself as morally responsible. Both Fischer and Ravizza's semicompatibilism and Mele's history-sensitive compatibilism view moral responsibility as a historical phenomenon. Thus, an agent needs to have a certain history in order for her to be morally responsible. The main difference in their historical accounts is that Mele rejects Fischer and Ravizza's subjective thesis.

As opposed to the indirect argument for incompatibilism that rejects the possibility for moral responsibility, we have reached a position where we can quite confidently describe circumstances in which we are morally responsible. Or at the least, we can give reasons why an agent is not morally responsible for her actions with reference to Fischer and Ravizza's beefed up conditions for an agent to act from an MRR mechanism combined with Mele's constraints on when the MRR mechanism is not an agent's own. However, this is not the end of the matter. As I noticed at the end of section 2, there may be one last avenue for the incompatibilist to pursue. Recall Jones from section 2. Suppose that in a determined world goddess Diana creates a zygote, which is thirty years

later Jones.² At the time of the creation, Diana combines her supreme knowledge of the laws of nature and the atoms that constitute the zygote so that Jones will exactly thirty years later deliberate that it is best to go the game. On the basis of that deliberation and fulfilling all the sufficient conditions for autonomous agency, Jones chooses to go to the game. Compare Jones with John who is just like Jones but lived his life in a normal deterministic world and goes to the game at the relevant time just like Jones. Can we contend that John is free and morally responsible while holding that Jones is not? (Ibid., 188—9.) Mele constructs a zygote argument from the basis of that question. (1) Because of the way Jones's zygote was created in a deterministic universe, he is not free and morally responsible. (2) With respect to free agency and moral responsibility, there is no significant difference between the way Jones's zygote comes to exist and the way normal zygote comes to exist in a deterministic universe. (3) Therefore, determinism rules out moral responsibility and free action. (Ibid., 189.)

The zygote argument dispels any concerns about unsheddable values or the agent's control over her mental lives being bypassed. Further, nothing precludes that the mechanism issuing in going to the game would be Jones's own. That is, nothing in the story suggests that Jones could not see himself as an agent and apt target for the reactive attitudes. Those conceptions of his are appropriately based on evidence. So what would take the compatibilist out from this very troubling corner? Note that the zygote argument is quite similar to the source argument for incompatibilism. Both of them argue that determinism entails that moral responsibility is completely out of reach. As a part of our compatibilist project, we have shown the way in which compelled agents differ from agents in a normal deterministic world. However, in the case of the zygote argument we may need to show, instead of the falsity of premise 2, that premise 1 is not true. Moreover, premise 2 of the source incompatibilist argument, if determinism obtains, no one is the ultimate source of her actions, is an analytic truth, given the description of determinism. (McKenna, 2009.) Therefore, the compatibilist needs to argue for the falsity of premise 1.

We must put off an adequate analysis of these arguments for some future paper but we can make some general observations. Although far-fetched, the zygote argument is an intriguing representation of how determinism poses a trouble for moral responsibility. Jones's going to the game is purely determined by forces beyond his control. Further, it seems that the incompatibilist can always fulfill the compatibilist conditions for moral responsibility by some responsibility-undermining way. Whether it is a goddess creating a zygote and a whole universe or a group of neurologists constructing an agent's psychological structure in a compatibilist-friendly fashion, there is still a strong intuitive pull not to hold these agents morally responsible for the pertinent conduct. Of course, there is a solution and we can always go ahistorical. However, that seems even

² This example is a direct adaptation of that Mele presents in "Free Will and Luck" (2006).

less tempting than conceding that the manipulated agents are morally responsible. This may be where the compatibilist is left. That is, Jones is responsible for going to the game.

References

Fischer, John Martin & Ravizza, Mark (1998), *Responsibility and Control: An Essay on Moral Responsibility*, Cambridge: Cambridge University Press.

Frankfurt, Harry G. (1969), "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy* 66/23, 829–39.

McKenna, Michael (2009), "Compatibilism", Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/win2009/entries/compatibilism/>.

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