

**REASONS-RESPONSIVENESS, ALTERNATIVE POSSIBILITIES, AND MANIPULATION ARGUMENTS AGAINST COMPATIBILISM: REFLECTIONS ON JOHN MARTIN FISCHER'S *MY WAY*.**

DERK PEREBOOM

*The University of Vermont*

John Martin Fischer's theory of moral responsibility is one of the great compatibilisms in the history of philosophy, standing alongside those of Aristotle, David Hume and Harry Frankfurt, for example, and of these it is arguably the most thoroughly developed. This superb collection of essays, *My Way*, contains Fischer's most recent thoughts about the theory, and like all of his work on these issues, it demands the close attention of anyone working in the area.

**1. Is Reactivity "all of a piece"?**

Central to the account of moral responsibility Fischer has worked out together with Mark Ravizza are two factors. The first is that an agent's moral responsibility for an action is explained not by the availability of alternative possibilities, but instead by the action's having a causal history of a sort that allows her to be the source of the action in a specific way. Thus they opt for a *source* as opposed to a *leeway* position. On the leeway side, the intuition that an agent's moral responsibility for an action requires that she could have done otherwise has considerable force. This force is nicely expressed by what David Widerker calls the *W-defence*.<sup>1</sup> About an agent who breaks a promise, but could not have done otherwise he writes:

Still, since you, [Harry] Frankfurt, wish to hold him blameworthy for his decision to break his promise, tell me *what, in your opinion, should he have done instead?* Now, you cannot claim that he should not have decided to break the promise, since this was something that was not in his power to do. Hence, I do not see how you can hold Jones blameworthy for his decision to break the promise. (Widerker, p. 191)

Despite the strong intuitions the *W-defence* captures, Fischer thinks that examples of the kind devised by Frankfurt yield an effective challenge to the leeway position.<sup>2</sup> In those examples an agent considers performing some action, but an intervener is concerned that she will not actually come through. Thus if the agent were to show some sign that she will not or might not perform the action, the intervener would cause her to perform the action anyway. So an intervener, Black, might ensure that an agent, Jones, will perform an action, say, killing Smith, by implanting a device in her brain, which upon detecting that she will or might not do so, would cause her kill

<sup>1</sup>. David Widerker, 'Frankfurt's Attack on Alternative Possibilities', *Philosophical Perspectives*, 14 (2000), pp. 181–201.

<sup>2</sup>. Harry G. Frankfurt, 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66 (1969), pp. 829–39; *My Way*, pp. 38–62.

Smith nevertheless. In actual fact Jones kills Smith on her own, without the intervention taking place. The intuition that Frankfurt aims to generate is that Jones could be morally responsible for killing Smith, despite the fact that she could not have done otherwise.

The second factor is the idea that the key feature of human agency that undergirds moral responsibility is the capacity to regulate behaviour by reasons. In the view Fischer initially developed, required for moral responsibility is a particular sort of *responsiveness to reasons*. He first argued that a weak notion of reasons-responsiveness is what is needed:

An agent is *weakly reasons responsive* when a certain kind K of mechanism, which involves the agent's rational consideration of reasons relevant to the situation, issues in action, and in *at least some alternative circumstances* in which there are sufficient reasons for her to do otherwise than she actually does, she would be receptive to these reasons and would have chosen and done otherwise by the efficacy of the same deliberative mechanism that actually results in the action.<sup>3</sup>

In response to counterexamples in which the agent is weakly reasons-responsive, and yet acts in accord with a pattern of reasons so odd that it seems clear that the agent is not morally responsible, in *Responsibility and Control* Fischer and Ravizza required that the mechanism by which the agent acts be receptive to a pattern of reasons that is intuitively rational. But at the same time, they in effect continued to endorse weak reasons-reactivity, on the ground that agents can be morally responsible when they are receptive to sufficient reasons to do otherwise, but yet do not react to those reasons.<sup>4</sup>

At a key point in *Responsibility and Control*, Fischer and Ravizza consider a challenge to their account. Brown, in taking a drug, Plezu, is moderately receptive but weakly reactive to reasons, and is morally responsible on the moderate reasons-responsiveness criterion. But Brown objects as follows:

It is unfair to hold me morally responsible for taking Plezu (a drug). After all, although I am regularly receptive to reasons, I am only weakly reactive to reasons. Thus, whereas I would have responded to a very different incentive for doing otherwise, the mechanism on which I acted did not—and *could not*—have responded to the *actual* incentive to do otherwise. Given this, it is unfair to hold me morally responsible. (p. 73)

Fischer and Ravizza respond in a way that has been controversial on several fronts. They claim that “reactivity is all of a piece”, that is, “if an agent's mechanism reacts to some incentive to (say) do other than he actually does, this shows that the mechanism can react to *any* incentive to do otherwise”.<sup>5</sup> However, there are examples that cause concern for this claim. Here is a case devised by Alfred Mele. Fred has agoraphobia, and as a result he has not left the house for ten

<sup>3</sup>. ‘Responsiveness and Moral Responsibility’, in Ferdinand Schoeman (ed.), *Responsibility, Character, and the Emotions* (Cambridge University Press, 1987), pp. 81–106, at p. 88.

<sup>4</sup>. John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge University Press, 1998), pp. 65–71.

<sup>5</sup>. *Responsibility and Control*, p. 73.

years even though has been receptive to sufficient reasons to do so. One day, he is receptive to a sufficient reason to attend his daughter's wedding at the church next door. But he does not attend due to his agoraphobia. Suppose that in a nearby world his house is on fire, and his agoraphobia is outweighed by his fear of the fire. Reacting to the reason provided by the danger posed by the fire, he leaves the house, and then attends the wedding.<sup>6</sup> The mechanism whereby Fred, in the actual situation, stays at home rather than attending the wedding next door is thus weakly reasons-reactive. But it is unintuitive that the mechanism by which he decides to stay at home can react to *any* incentive to do otherwise just by virtue of being able to react to the reason resulting from the fire. (Mele doesn't explicitly use this example to challenge the claim that reactivity is all of a piece; Jon Bicknell does employ a similar example to challenge this claim).<sup>7</sup>

Fischer and Ravizza also consider how it might be that an agent (agent, not mechanism) is able to react to a sufficient reason to do otherwise in some possible situation while it is false that the mechanism by which he acts can react to any incentive to do otherwise. The example they adduce is this:

Imagine that the agent somehow gets considerably more energy or focus if he is presented with a strong reason to do otherwise, and it is only in virtue of these factors that he successfully reacts to the reason. There certainly can be cases like this, but it is natural to say that, when the agent acquires significantly more "energy or focus," this gives rise to a *different mechanism* from the actual mechanism. Our point is that, holding fixed the actual kind of mechanism, reactivity is all of a piece: if the mechanism can react to any reason to do otherwise, it can react to all such reasons. (*Responsibility and Control*, p. 74).

To preserve the claim that reactivity is all of a piece, Fischer and Ravizza say that when the agent acquires more energy or focus, a new mechanism is at play. (A similar move might be made in the case of Fred, the wedding, and the fire.) But here the individuation of mechanisms might well seem ad hoc. Would there be a distinct mechanism for each increase of energy or focus? Intuitively, it does not seem implausible that the mechanism on which the agent acts can survive changes in energy or focus. The more general concern here is one raised by Michael McKenna:<sup>8</sup> the individuation of mechanisms should not be driven by the needs of the theory of moral responsibility, but rather by informed independent reflection on the nature of these mechanisms—I would say by psychology.

All of these problems could be avoided, however, if Brown's complaint were answered differently. First, Fischer and Ravizza's response, it seems to me, involves thinking about the reasons-responsiveness criterion as an ersatz principle of alternative possibilities. Brown's challenge was: 'the mechanism on which I acted did not—and *could not*—have responded to the *actual* incentive to do otherwise.' The conviction that there is a successful Frankfurt-style

---

<sup>6</sup>. Alfred Mele, 'Reactive Attitudes, Reactivity, and Omissions', *Philosophy and Phenomenological Research*, 61 (2000), pp. 447–52.

<sup>7</sup>. Jon Bicknell, 'From the Compatibilist Perspective: An analysis of John Martin Fischer's account of moral responsibility', College Honors Thesis, College of Arts and Sciences, University of Vermont, 2003.

<sup>8</sup>. Michael McKenna, Review of John Martin Fischer and Mark Ravizza *Responsibility and Control: A Theory of Moral Responsibility*, *Journal of Philosophy*, 98 (2001), pp. 93–100.

example, and that therefore the (relevant) principle of alternative possibilities is false, could, and perhaps should, generate a different response. It would be: ‘That is correct. But by virtue of the mechanism on which you acted, you did in actual fact respond to the incentive to do as you did.’ More generally, when Widerker asks about Joe (in the context of setting out the W-defence), in view of the fact that he had no robust alternative possibility: What would you have him do? I think that McKenna has it right: one should admit that there is no good answer. But, armed with the conviction that there is a successful Frankfurt-style case, one should instead call attention to what Joe has actually done, and to the causal history by which his action came about.<sup>9</sup>

Deploying a reasons-responsiveness criterion as kind of an ersatz principle of alternative possibilities may initially be attractive—PAP and a reasons-responsiveness criterion are at least superficially similar, and there are strong intuitions expressed by the W-defence that remain unanswered on a source view. But a Frankfurt-defender should hold the line. If there is a successful Frankfurt-style case, then the intuitions expressed by the W-defence are mistaken. Moreover, taking what is arguably the leeway incompatibilist bait here leads, at best, to a dialectical stalemate for the compatibilist. Fischer thinks that Frankfurt-defenders can do better than dialectical stalemate. So it makes sense to avoid this sort of answer to the leeway incompatibilist. He would then avoid the controversial claim that reactivity is all of a piece, and also the one instance in the corpus where a mechanism is individuated in a way that seems driven by the theory of moral responsibility.

## 2. Reasons-Responsiveness without Weak Reasons-Reactivity

Furthermore, I suspect that even weak reasons-reactivity is not a necessary condition for moral responsibility, for there would seem to be situations in which agents are not weakly reasons-reactive but could still be morally responsible. Effective examples are provided by actions that realize something that is of paramount value to the agent, so that he would realize it at any cost, but nevertheless he has and is receptive to what is in fact a sufficient reason for not realizing it.<sup>10</sup> The irrationality of such an agent need not be of the sort that undermines moral responsibility. Suppose that someone comes to your door and wants to know whether you are lodging a certain person—and you are. You are so committed to telling the truth that you would do so under any circumstances, even if, for example, you knew that the person at the door intended to murder your guest, or even if you knew that he would destroy the whole world if you told the truth. It would seem that you could be morally responsible for your truth-telling despite (the relevant mechanism) not being weakly reasons-reactive. Such cases would appear to indicate that weak reasons-reactivity is not necessary for moral responsibility.

One might argue that in this case you are no more responsible than the agoraphobe who will not leave the house for any reason. But suppose that your commitment is a Kierkegaardian one—Kierkegaard envisions an agent continuously renewing a commitment from moment to

<sup>9</sup>. Michael McKenna, ‘Where Frankfurt and Strawson Meet’, *Midwest Studies in Philosophy*, 29 (2005), pp. 163–80.

<sup>10</sup>. Derk Pereboom, *Living Without Free Will* (Cambridge University Press, 2001), p. 109, n. 41; here I also respond to the suggestion that such agents are responsible only by way of a tracing principle.

moment. We might thus imagine that as you make the decision to tell the truth to the person at the door, you also renew your commitment to the truth-telling principle—‘yes,’ you say to yourself, ‘that is a principle I will stand by, come what may.’ We might also specify that you have the traits of constancy and resoluteness to the degree that in any situation that features what is in fact a sufficient reason to renounce the principle, you are receptive to this reason, but you would not renounce it. Your constantly renewed commitment to truth-telling is therefore not weakly reasons-reactive. It seems that you could be morally responsible in this situation, despite the fact that your commitment to the principle, and the action based on it, are not weakly reasons-reactive.

Additional support for the claim that an agent can be morally responsible and not weakly reactive is provided by the following consideration. Suppose that an agent A is committed—in the Kierkegaardian sense—to three religious prohibitions, P1, P2 and P3. There are no possible circumstances in which A will violate P1 (that is, act in a way that he believes will violate P1) despite the fact that there are possible circumstances that feature what are in fact sufficient reasons to violate the prohibition, and A is receptive to those reasons. The only possible circumstances in which A will violate P2 are those in which he believes that by doing so she will avoid violating P1, and the only possible circumstances in which A will violate P3 are those in which she believes that by doing so she will avoid violating P1 or P2. Fischer’s criterion of moral responsibility allows that A can be morally responsible for acting so as to avoid violating P2 and P3, since in these cases the mechanism will be weakly reasons-reactive, but not for acting so as to avoid violating P1, for there it will not weakly reasons-reactive. This result seems counterintuitive.

How can the view be amended so as to allow for moral responsibility without weak reasons-reactivity? As noted above, in response to certain counterexamples to his earlier weak reasons-responsiveness account, Fischer and Ravizza substituted moderate reasons receptivity, which includes receptivity to the right sort of pattern of reasons, but stayed with weak reasons-reactivity. McKenna then argued, persuasively to my mind, that similar sorts of counterexamples indicate that weak-reasons reactivity is too weak a condition.<sup>11</sup> He plausibly contended that the account should instead require moderate reasons-receptivity and weaker reasons-reactivity. The idea is that the mechanism by which the agent acts should also be reactive in accord with a pattern of reasons, but since agents can be morally responsible when they are receptive to sufficient reasons to do otherwise but yet do not act on those reasons, the condition on the rationality of the pattern in accord with which the agent is reactive need not be as stringent as it is for receptivity.

Here is a suggestion that I believe respects all the considerations now canvassed: replace the moderate reasons-receptivity and weak reasons-reactivity requirements with this condition: the mechanism by which the agent acts must be receptive to a pattern of reasons that renders the mechanism rational (or rational enough), and this mechanism must be reactive in accord with this pattern of reasons, but in such a way that allows for it not to be reactive to sufficient reasons to which it is receptive. The resulting account dispenses with the requirement that at least in

---

<sup>11</sup>. Michael McKenna, ‘Assessing Reasons-Responsive Compatibilism’, *International Journal of Philosophical Studies*, 8 (2000), pp. 89–114; and ‘Reasons-Reactivity and Incompatibilist Intuitions’, *Philosophical Explorations*, 8 (2005), pp. 131–43, at pp. 133–6.

some situation in which the mechanism is receptive to a sufficient reason to do otherwise, it is reactive to that mechanism. That it does so is justified, for as we have seen, there are cases in which the agent is intuitively morally responsible in which this requirement is not met.

### 3. A Challenge for Fischer's Response to the 'Dilemma Defence' against Frankfurt-Style Arguments

Some leeway incompatibilists have contended that a close examination of Frankfurt-style cases actually substantiates their position. For such cases involve some factor that the intervener's device is set up to detect that could have occurred but does not actually occur in the agent, such as *forming an intention* to do to otherwise. The possible occurrence of such a factor—a 'flicker of freedom,' to use Fischer's term—is then proposed as the alternative possibility required for moral responsibility.<sup>12</sup>

Fischer argues, however, that one can construct Frankfurt-style examples in which the intervener's device detects some factor that occurs prior to the formation of the intention, and, more importantly, some factor that is intuitively not as intimately related to the action itself. For instance, one might imagine that Jones will decide to kill Smith only if Jones blushes beforehand. Then Jones's failure to blush (by a certain time) might be the alternative possibility that would trigger the intervention that would cause her to kill Smith. Supposing that Jones acts without intervention, we might well have the intuition that she is morally responsible for killing Smith, even though she could not have done otherwise than to kill Smith, and even though she could not even have formed an alternative intention. She could have failed to blush, but Fischer argues that such a flicker is of no use to the libertarian, since it is not sufficiently *robust* to have a role in grounding the agent's moral responsibility.<sup>13</sup>

It might now seem that any alternative-possibilities condition on moral responsibility can be defeated by a Frankfurt-style example that employs a non-robust flicker of freedom. But this line of defence for such arguments has proven to be too quick. For it is challenged by an important objection to these sorts of arguments that was initially raised by Robert Kane and then systematically developed by Widerker and Carl Ginet.<sup>14</sup> The general form of the Kane/Widerker/Ginet objection is this: for any Frankfurt-style scenario, if causal determinism is assumed to hold in that scenario, the libertarian will not have and cannot be expected to have the intuition that the agent is morally responsible. If, on the other hand, libertarian indeterminism is presupposed, an effective Frankfurt-style example is unavailable, for any such case will fall to a dilemma. In Frankfurt-style examples the actual situation will feature a prior sign that signals the fact that intervention is not necessary (such as the blush in Fischer's example). If in the proposed case the sign causally determined the action, or if it were associated with some factor that did so,

<sup>12</sup>. John Martin Fischer, *The Metaphysics of Free Will* (Blackwell, 1994), pp. 134–47.

<sup>13</sup>. *The Metaphysics of Free Will*, pp. 131–59.

<sup>14</sup>. Robert Kane, *Free Will and Values*, (SUNY Press, 1985), p. 51; Robert Kane, *Through the Moral Maze* (Paragon House, 1994), pp. 142–4, 191–2; David Widerker, 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities', *The Philosophical Review*, 104 (1995), pp. 247–61; and cf. Carl Ginet, 'In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Arguments Convincing', *Philosophical Perspectives*, 10 (1996), pp. 403–17.

the intervener's predictive ability could be explained. However, then the libertarian would not have the intuition that the agent is morally responsible. If the relationship between the sign and the action were not causally deterministic in such ways, then the libertarian would object that the agent could have done otherwise despite the occurrence of the prior sign. Either way, some alternative-possibilities condition on moral responsibility would emerge intact.

It is in response to Fischer's blushing case that Widerker suggests a no-determinism-in-the-actual-sequence requirement for Frankfurt-style arguments. Since then, many defenders of Frankfurt have attempted to devise examples in which the actual sequence is not deterministic, and yet no alternative possibilities, or none of the robust sort, are available to the agent. But in his reply to Widerker, Fischer makes an intriguing claim about the dialectical structure of the debate. He contends that it has already been made plausible by the early Frankfurt-style cases—those not designed to meet the no-determinism-in-the-actual-sequence requirement—that if the agent is not morally responsible, this is not simply because she could not have done otherwise, and the plausibility of this result is not undermined if it turns out that the causal history of the action in these cases is deterministic:

I think that the examples make highly plausible the preliminary conclusion that if Jones is not morally responsible for his choice and action, this is not simply because he lacks alternative possibilities. After all, everything that has causal (or any other kind of) influence on Jones would be exactly the same, if we “subtracted” Black [the intervener] entirely from the scene. And Jones's moral responsibility would seem to be supervenient on what has an influence or impact on him in some way. So the relevant (preliminary) conclusion is, if Jones is not morally responsible for his choice and action, this is not simply because he lacks alternative possibilities. And it does not appear to beg the question to come to this conclusion, even if causal determinism obtains. (‘Recent Work on Moral Responsibility,’ *Ethics*, 110 (1999), pp. 93–139, at p. 113)

For this reason Fischer maintains that the no-determinism-in-the-actual-sequence requirement is inadequately motivated.

I agree that the early Frankfurt-style arguments substantially enliven the possibility that facts about an action's actual causal history, rather than alternative possibilities, are pivotal in explaining an agent's moral responsibility, and that all by itself this yields at least some reason for thinking that these facts indeed have this explanatory role. But exactly how decisive a reason do these early arguments provide for this conclusion? The answer might depend on one's initial position in the debate, as Ishtiyaque Haji and McKenna have also contended.<sup>15</sup> Perhaps a Frankfurt-style argument of this early sort provides a more decisive reason for someone who is initially a leeway compatibilist than for someone who is initially a leeway incompatibilist. Suppose that a Frankfurt-style example is proposed in which it turns out that the actual causal history might well be deterministic, but the determinism is not of a sort that is responsibility-

<sup>15</sup>. Ishtiyaque Haji and Michael McKenna, ‘Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities’, *Journal of Philosophy*, 101 (2004), pp. 299–314; see also Derk Pereboom, ‘Source Incompatibilism and Alternative Possibilities’, in Michael McKenna and David Widerker (eds.), *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities* (Ashgate, 2003), pp. 185–99).

undermining given compatibilist intuitions. This example could well provide an Ayer-inspired leeway compatibilist with a strong reason to abandon her view, in favour of, say, Fischer's or Frankfurt's kind of compatibilism, each of which features an actual causal history account of moral responsibility. But it is not so clear that it should convince a leeway incompatibilist or an uncommitted participant in the debate who is concerned that determinism all by itself precludes moral responsibility.

A key question here is whether, in the Frankfurt-style examples at issue, in the last analysis it is causal determinism that explains the absence of alternative possibilities, since the presence of the intervener cannot preclude them without causal determinism in the actual sequence. If so, then even an uncommitted observer, who is nonetheless concerned that determinism might rule out alternative possibilities, should not be swayed away from a leeway view by a Frankfurt-style example. Stewart Goetz has pressed this point, making a case for the claim that “with determinism in the actual sequence it is not the device that prevents Jones from making an alternative choice”, and the sense that the device has this role is an illusion.<sup>16</sup> In his reply to Goetz, Fischer agrees that in this sort of example, determinism in the actual sequence is one factor that explains the agent's lack of alternative possibilities. However, he first notes that there are actually two factors that explain this lack of alternative possibilities—one is the causally deterministic process of the actual sequence, and the other is the non-actual process involving the intervener. He then points out that we can consider these two ensuring conditions separately, and bracketing the causally deterministic process of the actual sequence, we can focus just on the non-actual process involving the intervener. That non-actual process ensures the action in question, and yet we have an intuition that an agent can be morally responsible despite the conditions of this process being in place; or at least we have the intuition that if the agent is not morally responsible, it is not because the intervener's presence rules out alternative possibilities. An agent might then be morally responsible despite conditions being in place that ensure the action—conditions that do not depend on any causal determinism in the actual sequence to rule out alternative possibilities. So the uncommitted participant has reason deriving from this example to believe that absence of alternative possibilities is compatible with moral responsibility:

In the Frankfurt-type scenarios, *two* causes make it the case that Jones is unable to choose otherwise at T2: the prior condition of the world (together with the laws of nature) and Black's counterfactual intervention. What these examples show is that the mere fact that Jones is unable to choose otherwise does not in itself establish that Jones is not morally responsible for his choice. This is because Black's counterfactual intervention is one of the factors that make it the case that Jones is unable to choose otherwise at T2, and yet it is irrelevant to the grounding of Jones' moral responsibility. Considering this factor (the counterfactual intervention), and bracketing any other factor that might make it the case that Jones is unable to choose otherwise at T2, it seems to me that Jones may well be morally responsible for his [choice]. The mere fact that he lacks alternative possibilities, then, cannot in itself be the

---

<sup>16</sup> Stewart Goetz, 'Frankfurt-Style Arguments and Begging the Question', *Midwest Studies in Philosophy*, 29 (2005), pp. 83–105, at p. 85; Stewart Goetz, 'Alternative Frankfurt-Style Counterexamples to the Principle of Alternative Possibilities', *Pacific Philosophical Quarterly*, 83 (2002), pp. 131–47.

reason that Jones is not morally responsible, if indeed he is not morally responsible. (*My Way*, pp. 199–200)

In his reply to Fischer, Goetz argues that in general, the absence of alternative possibilities can only be explained by causal determinism in the actual sequence.<sup>17</sup> Perhaps a good case can be made for this general claim, but, as Frankfurt and Fischer have pointed out, it seems that in Frankfurt-style cases actions can be ensured by conditions that differ from the causal process that obtains in the actual sequence. There are two ways of thinking about determinism: (a) as involving the claim that events are *merely entailed* by propositions that describe preceding conditions and the laws of nature, and (b) as involving the claim that events are *actually produced* by such preceding conditions and the laws of nature. A proposition describing the conditions of a suitably constructed Frankfurt-style case entails that the action will come about, and so the action will be determined by these conditions in sense (a). But since all of the crucial features of these ensuring conditions are not actually operative, but some are merely counterfactual, such a case seems not to feature causal determinism in sense (b). Terminologically, to my ear sense (b) is causal determinism, while sense (a) involves determination by preceding conditions that does not obviously amount to *causal* determinism. And in my Tax Evasion case, I think I've produced a Frankfurt-style case in which an action is ensured so that the preceding conditions determine the action (in sense (a)) without causally determining them (in sense (b)).<sup>18</sup> Perhaps Goetz can provide a reason to believe that determination in sense (a) is impossible without causal determination—sense (b)—but I haven't yet seen the case made.

Let me now suggest a different response to Fischer—which is indebted to Kane, Widerker, Ginet, and Goetz. Fischer's account relies on the existence of two separate sets of ensuring conditions, one of which involves the merely counterfactual intervention, while the other consists just in the actual causally deterministic process that results in the action. Allow, as Fischer does, that the actual causally deterministic process that results in the action indeed constitutes one set of ensuring conditions. Let us focus instead, as Fischer counsels, on the set of ensuring conditions that involve the counterfactual intervention. He claims that this is indeed a set of ensuring conditions, and, to be sure, the key to the example's being a successful Frankfurt-style case is that it must be. However—and this is the crucial point—the presence of the counterfactual intervention, all by itself, that is, independently of what happens conditionally on the blush occurring by the specified time, only *conditionally* ensures that the action will occur. For the intervener's ensuring the action is conditional on the blush not occurring by the specified time. This conditional assurance does not, all by itself, *unconditionally* ensure the action, for in the set-up of the case, it is also possible that the blush does occur by the specified time. What produces the action if the blush does occur is what happens in the actual causal sequence, which, as Widerker has pointed out, must be causally deterministic relative to the blush's occurring if the action's occurrence is to be guaranteed on this condition. Consequently, if we are focusing on the ensuring condition involving the counterfactual intervention, given that unconditional

<sup>17</sup>. 'Frankfurt-Style Arguments and Begging the Question', pp. 91–2.

<sup>18</sup>. Pereboom, *Living without Free Will*, pp. 18–22

assurance is at issue, causal determinism in the actual causal sequence cannot be set aside—for it is now *the entire Frankfurt-example set-up with a causally deterministic actual sequence relative to the blush occurring* that unconditionally guarantees the action.

Merely conditional assurance is not at issue. It is easy to come by—for example, we might specify that the intervention would also occur if Jones puts on a red shirt. Suppose then that the intervention does not occur, and Jones kills Smith on her own, by virtue of a causally deterministic process. One might then venture: if we focus on the ensuring condition that the intervention would definitely occur if Jones were to put on a red shirt, we can see that she can be morally responsible despite this fact. So an agent can be morally responsible despite the unavailability of alternative possibilities. But this argument is unsound. The force of a Frankfurt-style case lies in the fact that the action is *unconditionally* ensured while it is nevertheless intuitive that the agent is morally responsible. To achieve unconditional assurance in the path involving the counterfactual intervention, the entire Frankfurt-case set-up is required. In a prior-sign case like Fischer’s blush example, if the prior sign that the agent will perform the action on her own occurs, the link between the blush and the action must be causally deterministic. Thus while this sort of case may involve two distinct sets of unconditionally ensuring conditions, as Fischer points out, the problem is that each set features, crucially, an actual sequence that is causally deterministic in a sense that undermines the force of the example, at least for certain key audiences.

So a Frankfurt example like this might well not be especially effective in giving the uncommitted participant or the incompatibilist reason to reject the principle of alternative possibilities. This is not a fatal problem for Frankfurt-style cases generally, since there are those, like the Mele-Robb case, Tax Evasion, and McKenna’s example, that avoid causal determinism in the actual sequence.<sup>19</sup>

#### 4. Source Incompatibilism and the Manipulation Argument.

If Frankfurt-style cases are successful, both source incompatibilism and source compatibilism remain as live options. According to source compatibilism, compatibilism is true, and an agent’s moral responsibility for an action is to be explained not by the existence of alternative possibilities available to her, but rather by the action’s having a causal history of a sort that allows the agent to be the source of her action in a specific way. Fischer is an advocate of a view of this kind, and he is thus an opponent of source incompatibilism. While he noted the possibility of source incompatibilism early on,<sup>20</sup> he argues that “there is simply no good reason to suppose that causal determinism in itself (and apart from considerations pertaining to alternative

---

<sup>19</sup>. Alfred Mele and David Robb, ‘Rescuing Frankfurt-Style Cases’, *The Philosophical Review*, 107 (1998), pp. 97–112; Alfred Mele, *Free Will and Luck* (Oxford University Press, 2006), pp. 87-95; Derk Pereboom, ‘Alternative Possibilities and Causal Histories’, *Philosophical Perspectives*, 14 (2000), pp. 119-37; Pereboom, *Living without Free Will*, pp. 18-22; Pereboom, ‘Source Incompatibilism and Alternative Possibilities’; see also David Hunt, ‘Moral Responsibility and Buffered Alternatives’, *Midwest Studies in Philosophy*, 29 (2005), pp. 126–45; Michael McKenna, ‘Robustness, Control, and the Demand for Morally Significant Alternatives’, in McKenna and Widerker (eds.), *Freedom, Responsibility, and Agency*, pp. 201-17.

<sup>20</sup>. John Martin Fischer, ‘Responsibility and Control’, *Journal of Philosophy*, 89 (1982), pp. 24-40.

possibilities) vitiates our moral responsibility”.<sup>21</sup> Michael Della Rocca contends that this is a claim for which he has not argued, and in fact I believe Fischer can be challenged on this point.<sup>22</sup> True, one incompatibilist intuition that many seem naturally to have is that if we could not act otherwise than we actually do, then we could never refrain from the immoral actions we perform, and for this reason we are not blameworthy for them. However, another very powerful and common intuition is that if all of our behaviour was ‘in the cards’ before we were born, in the sense that things happened before we came to exist that, by way of a deterministic causal process, inevitably result in our behaviour, then we cannot legitimately be judged blameworthy for our wrongdoing. By this intuition, if causal factors existed before a criminal was born that, by way of a deterministic process, inevitably issue in his act of murder, then he cannot legitimately be judged blameworthy for his action. If all of our actions had this type of causal history, then it would seem that we lack the kind of control over our actions that moral responsibility requires.

I do not believe that in the dialectic of the debate one should expect Fischer to be moved much by this incompatibilist intuition *alone* to abandon his compatibilism. Rather, the best type of challenge to the compatibilist at this point develops the claim that an action’s being produced by a deterministic process that traces back to factors beyond the agent’s control, even when she satisfies all the conditions on moral responsibility specified by the prominent compatibilist theories, presents in principle no less of a threat to moral responsibility than does deterministic manipulation. My ‘four-case argument’ first of all develops examples that involve such manipulation, in which these compatibilist conditions on moral responsibility are met, and which elicit the intuition that the agent is not morally responsible.<sup>23</sup> But further, it sets out three such cases, each progressively more like a fourth scenario, one that the compatibilist might envision to be realistic, in which the agent is causally determined to act in a natural way. The challenge for the compatibilist is to point out a difference between the fourth, ordinary, scenario that shows why the agent can be morally responsible in the ordinary case, and not in one or more of the manipulation examples. My suggestion is that non-responsibility generalizes from at least one of the manipulation cases to the ordinary one.

In each of the four cases I set out, Professor Plum decides to kill and kills Ms. White for the sake of some personal advantage. This act of murder conforms to the prominent compatibilist conditions, which are designed to be sufficient for an agent’s moral responsibility when supplemented by some fairly uncontroversial additional necessary conditions. This act satisfies the various conditions proposed by Hume and his followers: it is caused by desires that flow from his “durable and constant” character, since for him egoistic reasons typically weigh very heavily—much too heavily as judged from the moral point of view, while the desire on which he acts is nevertheless not irresistible for him, and in this sense he is not constrained to act.<sup>24</sup> It fits the condition proposed by Frankfurt: Plum’s desire to murder White conforms to his second-

<sup>21</sup>. *The Metaphysics of Free Will*, p. 159; *My Way*, pp. 131, 201–2.

<sup>22</sup>. Michael Della Rocca, ‘Frankfurt, Fischer, and Flickers’, *Noûs*, 32 (1998), pp. 99–105.

<sup>23</sup>. Pereboom, ‘Determinism al Dente,’ *Noûs* 29 (1995), pp. 21–45; Pereboom, *Living Without Free Will*, pp. 110–25.

<sup>24</sup>. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, (Oxford University Press, 1978), pp. 319–412

order desires in the sense that he wills to murder her and wants to will to do so, and he wills this act of murder because he wants to will to do so.<sup>25</sup> The action meets the moderate reasons-responsiveness condition set out by Fischer and Ravizza: for instance, Plum's desires are modified by, and some of them arise from, his rational consideration of the reasons at issue, he is receptive to the relevant pattern of reasons, and if he knew that the bad consequences for himself that would result from killing White would be much more severe than they are actually likely to be, he would have refrained from killing her for this reason.<sup>26</sup> It also satisfies a condition proposed by Jay Wallace: Plum retains the general capacity to grasp, apply, and regulate his behaviour by moral reasons. For example, when egoistic reasons that count against acting morally are relatively weak, he will typically regulate his behaviour by moral reasons instead. These capacities even provide him with the ability to revise and develop his moral character over time.<sup>27</sup> Now, given that causal determinism is true, is it plausible that Plum is responsible for his action?

In a first type of counterexample (Case 1) to these prominent compatibilist conditions, the manipulation is local—it proceeds from moment to moment. One might imagine that the sophisticated neuroscientists manipulate Plum from moment to moment by radio-technology to act in such a way that these compatibilist conditions are met. The manipulation takes place at the neural level, so that at the mental level of description the agent does not differ from one who is not manipulated. Hereby the incompatibilist aims to elicit the intuition that Plum is not morally responsible, thus showing that the prominent compatibilist conditions are not sufficient for moral responsibility. Even in an ordinary deterministic situation the fact that the agent meets the compatibilist conditions will not be sufficient for moral responsibility. Furthermore, I claim that there is no difference between Case 1 and Case 4—the ordinary deterministic scenario in which the agent meets the compatibilist conditions on moral responsibility—that can explain why an agent is not responsible when locally manipulated but is responsible in Case 4.

Perhaps the best reply on the part of the compatibilist involves developing the claim that a locally manipulated agent in some sense lacks the metaphysical integrity required to be a morally responsible agent, perhaps that a morally responsible agent must be one whose character develops in the ordinary way over a significant period of time and whose actions proceed from this character and not from external local deterministic manipulation. But then we might imagine Case 2, a scenario in which Plum is deterministically programmed by the neuroscientists from the beginning of his life so that his character develops as it does, to all appearances in an ordinary way, and so that he inevitably makes a decision that meets the prominent compatibilist criteria for moral responsibility. Is Plum blameworthy for the crime? The incompatibilist again aims to elicit the intuition that he is not morally responsible. It would seem unprincipled to claim that here, by contrast with the local manipulation example, Plum can now be morally responsible just because the length of time between the programming and the action is great enough. Whether the programming takes place two seconds or thirty years before the action seems irrelevant to the question of moral responsibility. Here again the incompatibilist claims that the

<sup>25</sup>. Frankfurt, 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, 68 (1971), pp. 5–20.

<sup>26</sup>. Fischer and Ravizza, *Responsibility and Control*, pp. 69–82.

<sup>27</sup>. R. Jay Wallace, *Responsibility and the Moral Sentiments* (Harvard University Press, 1994), pp. 51–83.

absence of responsibility generalizes to Case 4. (Case 3 is one in which Plum is causally determined by the training practices of his community to have the relevant sort of agency.)

By this argument, Plum's exemption from responsibility in Case 1 generalizes to his exemption from responsibility in Case 4. Notice that this argument is not a sorites. Its force does not depend on producing a series of cases, each of which is similar to its predecessor, and then arguing that since the first has some general feature, one must conclude that the last does so as well because each of the successive pairs of cases are different only in some small degree of that type of general feature. A series of similar cases is indeed important to the argument. But its strength derives from the fact that between each successive pair of cases there is no divergence at all in factors that could plausibly make a difference for moral responsibility, and that we are therefore forced to conclude that all four cases exhibit the same kind of responsibility-undermining feature.

Fischer has developed a challenge to this four-case manipulation argument.<sup>28</sup> He contends that Plum *is* morally responsible in Cases 1 and 2, and that our intuition that he is not morally responsible stems from the correct sense that he is not *blameworthy*:

In my view, further conditions need to be added to mere guidance control to get to blameworthiness; these conditions may have to do with the circumstances under which one's values, beliefs, desires, and dispositions were created and sustained, one's physical and economic status, and so forth. Professor Plum, it seems to me, is not blameworthy, even though he is morally responsible. That he is not blameworthy is a function of the circumstances of the creation of his values, character, desires, and so forth. But there is no reason to suppose that anything like such unusual circumstances obtain merely in virtue of the truth of causal determinism. Thus, I see no impediment to saying that Plum can be blameworthy for killing Mrs. White in Case 4. Note that there is no difference with respect to the minimal control conditions for moral responsibility in Cases 1 through 4—the threshold is achieved in all the cases. But there are . . . wide disparities in the conditions for blameworthiness. (*My Way*, p. 233)

I agree that there are cases in which an agent is morally responsible without being blameworthy—when she is praiseworthy for having performed a morally exemplary action, or when she performs an action that is morally indifferent. But could Plum, whose action is immoral, be morally responsible without being blameworthy? In my view, an agent's being blameworthy for an action is in fact entailed by his being morally responsible for it in the sense at issue in the debate, together with his understanding that the action was in fact morally wrong. This is because for an agent to be morally responsible for an action in the sense at issue is for it to belong to him in such a way that he would deserve blame if he understood that it was morally wrong, and he would deserve credit or perhaps praise if he understood that it was morally exemplary, supposing that this desert is basic in the sense that the agent would deserve the blame or credit just because he has performed the action (given understanding of its moral status), and not by virtue of consequentialist considerations. Assuming this characterization, and Plum's

---

<sup>28</sup>. *My Way*, pp. 230–4.

understanding that killing White is morally wrong, he could not be morally responsible for committing this murder without also being blameworthy for it.

But there are other senses of ‘moral responsibility’. One might say that an agent is morally responsible when it is legitimate to demand of her that she explain how her decisions accord with morality, and that she evaluate critically what her decisions indicate about her moral character.<sup>29</sup> Making these demands of agents might be justified by its effectiveness in improving the agent morally—we humans are indeed susceptible to causal influence by challenge and counsel of this kind. While this ‘legitimately calling to moral account’ notion may be a *bona fide* sense of moral responsibility, it is not the one at issue in the free will debate. For incompatibilists would not find our being morally responsible in this sense to be even *prima facie* incompatible with determinism. The notion that incompatibilists do claim to be incompatible with determinism is rather the one defined in terms of basic desert.

If Plum is morally responsible in Case 1 and Case 2 in the ‘basic desert’ sense that is at issue, then given his understanding that his action is morally wrong, it is entailed that he is blameworthy. An intuition that Plum can be morally responsible without being blameworthy might be explained by the possibility that he might be responsible in some other sense while not being blameworthy (still, there might be an analogue of blameworthiness that correctly applies to him—for instance, he might be legitimately admonished).<sup>30</sup> But it is not responsibility in this sense that incompatibilists have thought to be at odds with determinism. If one would want to deny this entailment claim, it would be valuable to have an example in which it is clear that an agent performs an action that is morally wrong, she understands that it is morally wrong, she is morally responsible for that action in the sense at issue in the debate, and nevertheless she is not blameworthy for performing it. Then, once the example is in place, we can try to figure out whether Plum’s situation in the examples at issue is relevantly analogous.

That said, I think that the ‘legitimately calling to moral account’ sense of moral responsibility might well be the most significant one that we in fact instantiate given the best philosophical arguments and scientific theories we have about the nature of the physical world, and that Plum in these cases might be responsible in this sense. Moreover, I think that the notion of moderate reasons-responsiveness (perhaps revised as suggested) yields the most promising account of this sort of responsibility. So I suspect I can agree with Fischer on the following claim: moderate reasons-responsiveness provides the most promising account of what might be the most significant kind of moral responsibility we actually have.<sup>31</sup>

---

<sup>29</sup>. This characterization is inspired by Arthur Kuflik, in conversation, and by Hilary Bok, *Freedom and Responsibility* (Princeton University Press, 1998), p. 152.

<sup>30</sup>. Pereboom, *Living Without Free Will*, pp. 156–7.

<sup>31</sup>. Thanks to Michael McKenna, John Fischer, Seth Shabo, Louis deRosset and David Christensen for valuable comments and discussion.