

JOHN MARTIN FISCHER

# RESPONSIBILITY AND MANIPULATION

(Received 30 September 2003; accepted 1 October 2003)

ABSTRACT. I address various critiques of the approach to moral responsibility sketched in previous work by Ravizza and Fischer. I especially focus on the key issues pertaining to manipulation.

KEY WORDS: alternative possibilities, causal determinism, free will, manipulation, moral responsibility

## I. INTRODUCTION

A compatibilist about causal determinism and moral responsibility wishes to say that the mere fact that the behavior in question is the product of a causally deterministic sequence does not imply that the agent cannot legitimately be held morally responsible for it. At the same time, the compatibilist typically is willing to concede that certain sorts of causal sequences undermine moral responsibility. Certain kinds of "manipulation" that bypass or somehow supercede or fundamentally distort the human capacity for practical reasoning are salient examples of responsibility-undermining factors. Now the challenge is to explain the difference between those sequences that undermine responsibility and those that are consistent with it (and, indeed, confer it). If it is not true that all causal sequences are created equal, how do we distinguish them?

This is a challenge I have sought to address head-on.[1] It is not an easy task, and my preliminary attempts have not elicited unanimous agreement. Below I shall discuss some of the most powerful critical discussions. I wish to begin by thanking my critics for their patient and sympathetic reading of my views, and for their penetrating critiques, from which I have learned much.

[1] John Martin Fischer and Mark Ravizza, "Responsibility and History," in Peter French, Theodore E. Uehling, Jr. and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy 19: Philosophical Naturalism* (Notre Dame: University of Notre Dame Press, 1994), pp. 430–451; and *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998), especially pp. 207–239.

*The Journal of Ethics* **8:** 145–177, 2004.

## II. A THEORY OF MORAL RESPONSIBILITY

I shall offer a brief sketch of my approach to moral responsibility in order better to understand the various critiques.[2] The theory has various major components. First, I argue that moral responsibility does not require genuine access to metaphysically open alternative possibilities; thus, causal determinism does not threaten moral responsibility (simply) in virtue of eliminating such access to alternative possibilities. In the course of elaborating this argument, I distinguish between two kinds of control. Regulative control involves genuine access to alternative possibilities, whereas guidance control does not. I thus contend that moral responsibility implies guidance control, but not regulative control. Guidance control is the "freedom-relevant" (as opposed, say, to "epistemic") condition that is both necessary and sufficient for moral responsibility.

I go on to argue that an agent exhibits guidance control of his behavior insofar as it issues from his own, moderately reasons-responsive mechanism. I presuppose a distinction between the kind of mechanism that actually results in the behavior and other sorts of mechanisms. Given that the actual mechanism is identified, it must be the agent's own, and it must be appropriately sensitive to reasons (including moral reasons).

I (and my co-author) elaborate the various components of guidance control at some length in *Responsibility and Control: A Theory of Moral Responsibility*. I offer only the briefest of sketches here. One has control of one's behavior at least in part in virtue of having *taken control* of the mechanisms that produce it. One takes control by *taking responsibility*. Taking responsibility involves three elements. First, the agent must see that his choices have certain effects in the world – that is, he must see himself as the source of consequences in the world (in certain circumstances). Second, the individual must see that he is a fair target for the reactive attitudes as a result of how he affects the world. Third, the views specified in the first two conditions – that the individual can affect the external world in certain characteristic ways through his choices, and that he can be fairly praised and/or blamed for so exercising his agency – must be based on his evidence in an appropriate way.[3]

In an earlier work, *The Metaphysics of Free Will: An Essay on Control*, I presented a preliminary sketch of the account of guidance control.[4] In the

[2] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.

[3] My co-author and I develop the account of taking responsibility at greater lengh in Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 207–239.

[4] John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control* (Oxford: Blackwell Publishers, 1994).

early presentation, I included only the reasons-sensitivity component, and I explicitly pointed out that this was a mere adumbration of a fuller account to be presented later. Specifically, I noted that the relevant sort of reasons-responsiveness could be induced by manipulation (or other responsibility-undermining factors), and that I would address this problem in future work. The added component of mechanism-ownership is an innovation in the account of guidance control presented in *Responsibility and Control: A Theory of Moral Responsibility*, and I (and my coauthor) suggest there that it can help with the problems of manipulation.

The intuition is simple. The mechanism that issues in behavior (or, more broadly, the way the behavior is produced) can be reasons-responsive, but this sensitivity, or significant features of it, could have been induced externally (by clandestine manipulation, hypnosis, subliminal advertising, brainwashing, and so forth). So reasons-sensitivity is not enough for moral responsibility. The reasons-responsiveness itself cannot have been put in place in ways that bypass or supercede the agent – the mechanisms that issue in one's behavior must be *one's own*.

## III. STUMP'S CRITIQUE

1. *Stump's first critique.* In various papers, Eleonore Stump has offered vigorous criticisms of elements of the overall account of moral responsibility I (and my co-author) have presented.[5] In her recent paper, "Control and Causal Determinism," she offers two criticisms I wish to discuss here.[6] She first points out that my co-author and I simply assume that there can be reasons (and agents can have reasons) in a causally deterministic world.

---

[5] See, for example, Eleonore Stump, "Control and Causal Determinism," in Sarah Buss and Lee Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge: MIT Press, 2002), pp. 33–60.

[6] In "Control and Causal Determinism," Stump also develops a critique of the criticism of the Direct Argument for Incompatibilism offered by Ravizza and me. The Direct Argument purports to show that causal determinism rules out moral responsibility, quite apart from considerations pertaining to alternative possibilities. It employs a modal principle that alleges that nonresponsibility can be transferred in a characteristic way. Ravizza and I criticize this argument in Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 151–169. Stump's criticisms are on pages 38–46; she offers related criticisms in, Eleonore Stump, "The Direct Argument for Incompatibilism," *Philosophy and Phenomenological Research* 61 (2000), pp. 459–466. Stump's paper is a contribution to a book symposium on *Responsibility and Control: A Theory of Moral Responsibility*; Ravizza and I reply to Stump in John Martin Fischer and Mark Ravizza, "Reply to Critics," *Philosophy and Phenomenological Research* 61 (2000), pp. 477–480.

Actually, Stump frames her critique here in terms of "tracking reasons." That is, she contends that Ravizza and I simply assume that agents can "track reasons" even in a causally deterministic work, but that we offer no argument for our claim. I suppose that the best way to interpret Stump is as follows: although we offer an account of the specific sort of "tracking reasons" that is involved in moral responsibility – moderate reasons-responsivenes – and we argue that this sort of tracking is entirely consistent with moral responsibility, if *any* kind is, we do not offer any sort of answer to the more fundamental question of whether *any* kind of tracking of reasons is consistent with casual determinism. Stump points out that the more fundamental idea is "crucial to our case" for compatibilism, and she goes on to say, "Without some way of supporting it, Fischer and Ravizza do not have an *argument* for their compatibilism."[7]

In supporting her criticism, Stump invokes the authority of such eminent philosophers as Patricia Churchland and Richard Rorty. She cites Churchland as follows:

> Boiled down to essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing, fighting, and reproducing. The principal chore of the nervous system is to get the body parts where they should be in order that the organism may survive ... Truth, whatever that is, definitely takes the hindmost.[8]

And Rorty says:

> The idea that one species of organism is, unlike all the others, oriented not just toward its own increased prosperity but toward Truth, is as un-Darwinian as the idea that every human being has a built-in moral compass ...[9]

I find this criticism perplexing. Yes, my co-author and I did simply assume that there is nothing in the very nature of causal determinism or reasons that would preclude agents in a causally deterministic world from having reasons or tracking reasons (quite apart from any particular account of reasons-tracking). But this is not an implausible position, and it has been argued for (convincingly, we should have thought) by various philosophers.[10]

[7] Stump, "Control and Causal Determinism," p. 38.

[8] Stump, "Control and Causal Determinism," p. 36; the quotation from Churchland is from Patricia Churchland, "Epistemology in the Age of Neuroscience," *The Journal of Philosophy* 84 (1987), pp. 548–549.

[9] Stump, "Control and Causal Determinism," pp. 36 and 38; the quotation from Rorty is from Richard Rorty, "Untruth and Consequences," *New Republic* (July 31, 1995), p. 36.

[10] See, for example, Daniel Dennett, "Intentional Systems," *The Journal of Philosophy* 68 (1971), pp. 87–106; and "Mechanism and Responsibility," in T. Honderich (ed.), *Essays on Freedom of Action* (London: Routlege and Kegan Paul, 1973), pp. 159–184. See also Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge: MIT Press, 1984); and Daniel Dennett, *Freedom Evolves* (New York: Viking, 2003).

Further, our overall theory has various parts; we offer arguments seeking to establish (or render plausible) various of these elements. Does one not have an *argument* for a contentious philosophical position unless one offers explicit justifications for *every* element of it? For all of its background assumptions and presuppositions? For the methodology one employs in seeking to support it? I would suggest that the methodological views suggested by Stump's critique are impossibly demanding.

Turning to the views of the luminaries, I simply do not see how they are relevant. In developing our account of moral responsibility, we do employ the notion of "reason." But we do not present a specific account of reasons – their ontological status or logic. Our goal was to present at least the rudiments of a systematic theory of moral responsibility – one that could be employed (perhaps *mutatis mutandis*) by proponents of a broad range of particular accounts of reasons. We would hope that the acceptability of a general theory of moral responsibility would not hinge on the viability of any particular (contentious) account of reasons.

So we were rather vague about reasons. We certainly did not say, nor, as far as I can tell, are we committed to the idea that reasons presuppose that there is anything like "Truth," with a capital "T," or that human beings are uniquely "oriented" to "It" (whatever "It" is). An organism – any organism – can have reasons insofar as he or she can have interests or a "stake" in something. But there are various particular ways of unpacking the concept of reasons (or perhaps their nature or essence), as well as their logic. Nothing in our theory requires us to say that there is some objectionably or problematically objective notion of truth, nor does it require that we bestow hegemony on human beings. Perhaps (for all we have said or are committed to, simply in virtue of offering a theory of moral responsibility) reasons are factors that make (or are taken to make) success in the four F's more likely, or they are the mental states that constitute awareness of such factors, or ... A theory of moral responsibility is, after all, more abstract than a theory of reasons; and certainly it is more abstract than a "first-order" theory in ethics (such as utilitarianism or Kantianism, and so forth). So I conclude that Stump's critique here is, if I may put it this way, a bit "reproduced-up."[11]

2. *Stump's second critique.* Stump's second critique is more probing. She argues that our new account of moral responsibility cannot adequately handle various manipulation cases, even in spite of the new element of

[11] For a similar conclusion, put in a considerably more genteel fashion, see Harry Frankfurt, "Reply to Eleonore Stump," in Buss and Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge: MIT Press, 2002), pp. 61–63.

mechanism ownership. Indeed, Stump suggests that it is precisely this element that yields unintuitive results in a range of manipulation cases.

It will be helpful to have before us the details of Stump's presentation. She begins:

A person who is being manipulated by someone else can meet [Fischer and Ravizza's] conditions for acting on a mechanism that is his own and also suitably reasons-responsive. Consequently, a manipulated person can count as morally responsible on their account of moral responsibility.

To see that this is so, consider Robert Heinlien's *The Puppetmasters*. In the story, an alien race of intelligent creatures wants to conquer the Earth. Part of the alien plan for invasion includes a covert operation in which individual aliens take over particular human beings without being detected. When an alien "master" takes over a human being, the human being (say, Sam) has within himself not only his own consciousness but the master's as well. The master can control Sam's consciousness; he can make Sam's mind blank, he can suppress or even eradicate some affect of Sam's, or he can introduce thoughts and desires into Sam's consciousness. Most of the time, however, the master leaves Sam's consciousness alone but simply takes it off-line. That is, Sam's consciousness runs pretty much as always, but it has no effect on Sam's behavior; the master's consciousness causes Sam to do whatever he does. The master controls Sam indirectly, by controlling Sam's thoughts and desires and then letting Sam's consciousness produce Sam's behavior.

Since it is crucial to the alien plan that their taking over human beings be undetected in the early stages of the invasion, they are careful to make the behavior of people like Sam correspond to the behavior Sam would normally have engaged in had he not been infected with the alien. So when, under the control of the alien, Sam does A, it is also true that if there had been reason sufficient for Sam in his uninfected state to do not-A, the alien would have brought it about that Sam in his infected state did not-A. In this case, then, Sam acts on a mechanism that meets Fischer and Ravizza's condition for being strongly reasons-responsive: "if [a certain kind of mechanism] K 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."[12]

Stump continues:

Suppose that we now rewrite Heinlein's story a little, in order to take account of Fischer and Ravizza's conditions for a mechanism's being an agent's own. Let it be the case that, after the alien has infected Sam and before he starts to manipulate Sam's reason, the alien has what is, in effect, a conversation with Sam. The alien may have no purpose for this conversation other than to amuse himself. But suppose that, for amusement or some other purpose, the alien wants to convince Sam that when Sam acts under the control of the alien, Sam is as much an agent and as suitable a candidate for the reactive attitudes of others as he ever was in his uninfected state.

The alien might, for example, put forward arguments for determinism and compatibilism that Sam finds extremely plausible. In consequence, Sam might come to believe that all the states of his mind and will are causally determined by factors outside himself and that, nonetheless, when he acts, determined in this way, he is incontrovertibly an agent and that it is perfectly appropriate for others to maintain the reactive attitudes toward him.

[12] Stump, "Control and Causal Determinism," pp. 47–48.

Next, the alien might argue to this effect: It can make no difference to our assessment of a person S whether the external factors determining the states of S's mind and will are animate or inanimate, intelligent or blind. Our assessment of S himself should remain the same regardless of whether or not the causes determining S include something sentient among them. Suppose that Sam finds this argument, too, very plausible. By this means, Sam, in the revised story, is brought to believe that, in acting on his mind and will as they are controlled by the alien, he *is* an agent and a suitable target for the reactive attitudes of others, just as he was in his uninfected state. These beliefs of Sam's will be false, but, of course, it is possible for human beings to reason themselves into very peculiar false beliefs ... Furthermore, these beliefs of Sam's will be founded on the evidence available to Sam, namely, what Sam knows and believes and the arguments of the alien which Sam accepts. ... In this way, then, Sam takes responsibility for the mechanism on which he acts when he is controlled by the alien, and so this mechanism counts as his own, on Fischer and Ravizza's account. Since this mechanism is also reasons-responsive in the way I described, Sam meets the Fischer and Ravizza conditions for moral responsibility when he is controlled by the alien. ... I think that the case of Sam and the puppetmaster is enough to show that Fischer and Ravizza's account has a serious problem in attempting to deal with manipulation ...[13]

Stump goes on to discuss two examples that Ravizza and I presented. She contends that her analysis further elaborates the problem suggested by the Puppetmaster case:

Here is the first case [Fischer and Ravizza's Judith I]:

A scientist secretly implanted a mechanism in Judith's brain (let us say, a few days ago). Employing this mechanism, the scientist electronically stimulates Judith's brain in such a way as to create what will be a literally irresistible urge to punch her best friend, Jane, the next time she sees Jane. When Judith meets Jane at a local coffeehouse, Judith experiences this sort of urge, and does indeed punch Jane.[14]

Our intuitive response to this case is to think that Judith is not responsible for punching Jane. Fischer and Ravizza think that their account can support this intuition ... But it is not difficult to flesh out *Judith I* in such a way that our intuition about the case remains the same, and yet Fischer and Ravizza's account on longer supports that intuition. We can easily assimilate *Judith I* to the sort of story in the revised version of Heinlein's *Puppetmasters*. In that case, the mechanism on which Judith acts in *Judith I* is the mind of the manipulator operating on her brain. As in the case of *Puppetmasters*, we can also suppose that the mechanism is suitably responsive to reasons that both Judith and the manipulator recognize as reasons for Judith, so that the mechanism is even strongly reasons-responsive. Finally, we can imagine that Judith comes to believe that she is an agent and the appropriate target of the reactive attitudes when she is controlled in this way by the manipulator.

Consequently, contrary to what Fischer and Ravizza suppose, a person such as Judith who acts on an irresistible desire produced in her by a manipulator can still meet the Fischer and Ravizza conditions for moral responsibility. She can act on a mechanism that is her own, in virtue of the fact that she has taken responsibility for it, and that mechanism can be

[13] Stump, "Control and Causal Determinism," pp. 49–50. She goes on to consider even more complex cases, but I think the reply I shall give in the text applies to all of her cases.

[14] The example comes from Fischer and Ravizza, *Responsibility and Control*, p. 231.

suitably reasons-responsive, because the manipulator manipulates his victim in a way that tracks reasons for the victim.[15]

Stump goes on to consider another case, *Judith II*, but we shall focus on her analysis of *Judith I* and her Sam and the *Puppetmasters* case. I pause to note that no less an authority than Harry Frankfurt is in agreement with Stump's criticism:

> Fischer and Ravizza seek to insulate their account of moral responsibility against the possibility that someone who is manipulated by another person might be wrongly held to be morally responsible for what he does. It seems to me that Stump is correct in her claim that their attempt to accomplish this insulation is unsuccessful. Her discussions of the examples involving Sam and Judith show effectively that even an agent who is being manipulated in ways that undermine moral responsibility can, according to the criteria Fischer and Ravizza provide, act on a mechanism that is both suitably reasons-responsive and the agent's own. Thus she shows that their criteria do not satisfactorily identify the conditions upon which moral responsibility depends.[16]

Of course, I hate to spoil the party. But I do not think that Stump's criticism is on target. Note that Stump contends, "... the mechanism on which Judith acts in *Judith I* is the mind of the manipulator operating on her brain." She goes on to write, "As in the case of *Puppetmasters*, we can also suppose that that mechanism is suitably responsive to reasons ..." Why does Stump suggest that in the case of *Puppetmasters*, Sam's mechanism is reasons-responsive? Recall that Stump argues:

> Sam acts on a mechanism that meets Fischer and Ravizza's conditions for being strongly reasons-responsive: "if [a certain kind of mechanism] K 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."[17]

And this is because:

> ... when, under the control of the alien, Sam does A, it is also true that if there had been reason sufficient for Sam in his uninfected state to do not-A, the alien would have brought it about that Sam in his infected state did not-A.[18]

Well, if you take the relevant mechanism (on which the agents in question act) to be individuated as broadly as "the mind of the manipulator acting on her brain," then of course it will turn out that the mechanism in question is in the specified way reasons-responsive. Similarly in the case of Sam, and in *any* manipulation case, if the mechanism is individuated as broadly as "manipulation by an external source," then, of course, the

15 Stump, "Control and Causal Determinism," pp. 50–51.

16 Frankfurt, "Reply to Eleonore Stump," p. 61.

17 Stump, "Control and Causal Determinism," p. 48.

18 Stump, "Control and Causal Determinism," p. 48.

mechanism will turn out to be reasons-responsive. This is because, no matter how thoroughly and effectively the external source actually manipulates the agent to do *X*, under other circumstances the source could have manipulated the agent in a different way to cause the agent to do *not-X*.

I should have thought that this very basic point could be seen to apply even to the simplest cases of manipulation. That is, it should be evident that, in order to render the Fischer–Ravizza account of manipulation cases even minimally plausible, we are not thinking of the relevant mechanisms as individuated so broadly as, for example, "manipulation by an external source." Rather, the mechanism is something like, "manipulation of this specific sort," where the sort in question is some is specified at least in part in terms of neurophysiology.

It is hard to see how there could be any confusion about how my co-author and I intend the account to work in this specific respect. For example, we say about Judith I:

> ... Here it is evident that Judith should *not* be held morally responsible for punching Jane. On our approach to moral responsibility, there are two distinct reasons why this is so. First, the mechanism leading to the action is not moderately reasons-responsive; by hypothesis, given the kind of stimulation of the brain that actually takes place, Judith as an irresistible urge to strike Jane. Thus, Judith would strike Jane, no matter what kinds of reasons to refrain were present.[19]

The account of manipulation only works, if it works at all, if one holds fixed the actual kind of brain manipulation, when one holds fixed the kind of mechanism that actually operates. This point is simple and straightforward; if it is not accepted, then one can criticize the Fischer–Ravizza account of moral responsibility right from the start, employing the examples we originally employed; the point does not pertain at all to the account of "one's own mechanism" or "taking responsibility," and no complicated examples such as Sam and the Puppetmasters need be invoked.

Consider, also, the Fischer–Ravizza discussion of "irresistible desires" or "compulsions." Obviously, there need be no external manipulation or induction for an agent to experience an irresistible urge; we might call this sort of urge a "compulsion." Now if the mechanism in question is individuated as broadly as "practical reasoning" or "deliberation," then (say) practical reasoning that involves a compulsive desire will be perfectly reasons-responsive. In order for our account even to get off the ground here, we must be considering the relevant mechanisms as individuated more narrowly. And we say, when first discussing such examples:

---

[19] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 231–232.

Consider, then, the mechanism, "deliberation involving an irresistible desire." Whereas this mechanism is temporally intrinsic, it is also reasons-responsive: there is a possible scenario in which Jim acts on this kind of mechanism and refrains from taking the drug. In this scenario, Jim has an irresistible urge to refrain from taking the drug. This shows that neither "deliberation involving an irresistible desire for the drug" [because it is not temporally intrinsic] nor "deliberation involving an irresistible desire" is the relevant mechanism (if the theory of responsibility is to achieve an adequate "fit" with our intuitive judgments).

When Jim acts on an irresistible urge to take the drug, there is some physical process of kind *P* taking place in his central nervous system. When a person undergoes this kind of physical process, we say that his urge is literally irresistible. And we believe that what underlies our intuitive claim that Jim is not morally responsible for taking the drug is that the relevant kind of mechanism issuing in Jim's taking the drug is of physical kind *P*, and that a mechanism of kind *P* is not reasons-responsive.[20]

Stump's critique, then, is off the mark because it employs an overly broad notion of mechanism-individuation, contrary to the explicit development of the theory. Further, despite Stump's suggestion that the problems come from the new component of the theory that specifies how agent's make the springs of their action their own by taking responsibility for them, the alleged problems come entirely from the original component of guidance control – reasons-responsiveness.

Now it might be noted that so far I have simply pointed out that the Fischer–Ravizza view depends on a certain notion of mechanism-individuation – one quite different from the one adopted, for the sake of her criticism, by Stump. But this is not yet to say that our notion of mechanism-individuation is the "correct" one. Perhaps the problem is not quite the one identified by Stump, but a problem nevertheless. I fully admit that this element of the overall account of moral responsibility is left to some degree vague, and that it is therefore at least to some degree problematic. It is thus entirely fair to point to problems that arise out of this vagueness. Stump's critique helpfully points to some of the commitments of our theory, and challenges us to say more about them. I shall return to these issues in Section VII.

## IV. PEREBOOM'S CRITIQUE

In his book, *Living without Free Will*, Derk Pereboom presents what he takes to be a problem for *any* compatibilist account of moral responsibility.[21] Pereboom starts with a case in which he believes that anyone

[20] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, p. 48.

[21] Derk Pereboom, *Living without Free Will* (Cambridge: Cambridge University Press, 2001), pp. 110–126.

would say that the agent is not morally responsible. He then transforms that case, step by step, into a context of causal determinism. Pereboom's position is that the compatibilist cannot distinguish, in a principled way, between cases in which we would all agree that there is not moral responsibility and the context of causal determinism.

Here is the first case:

Case 1. Professor Plum was created by neuroscientists, who can manipulate him directly through the use of radio-like technology, but he is as much like an ordinary human being as is possible, given this history. Suppose these neuroscientists "locally" manipulate him to undertake the process of reasoning by which his desires are brought about and modified – directly producing his every state from moment to moment. The neuroscientists manipulate him by, among other things, pushing a series of buttons just before he begins to reason about his situation, thereby causing his reasoning process to be rationally egoistic. Plum is not constrained to act in the sense that he does not act because of an irresistible desire – the neuroscientists do not provide him with an irresistible desire – and he does not think and act contrary to character since he is often manipulated to be rationally egoistic. His effective first-order desire to kill Ms. White conforms to his second-order desires. Plum's reasoning process exemplifies the various components of moderate reasons-responsiveness. He is receptive to the relevant pattern of reasons, and his reasoning process would have resulted in different choices in some situations in which the egoistic reasons were otherwise. At the same time, he is not exclusively rationally egoistic since he will typically regulate his behavior by moral reasons when the egoistic reasons are relatively weak – weaker than they are in the current situation.[22]

Pereboom's intuition is that Professor Plum is clearly not morally responsible in this case. He goes on to construct a case in which there is no local manipulation, but in which he believes that we will also agree that there is no moral responsibility:

Case 2. Plum is like an ordinary human being, except that he was created by neuroscientists, who, although they cannot control him directly, have programmed him to weigh reasons for action so that he is often but not exclusively rationally egoistic, with the result that in the circumstances in which he now finds himself, he is causally determined to undertake the moderately reasons-responsive process and to possess the set of first-and second-order desires that results in his killing Ms. White. He has the general ability to regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and accordingly he is causally determined to kill for these reasons. Nevertheless, he does not act because of an irresistible desire.[23]

Now Pereboom constructs a case in which the neuroscientists are replaced by parents, community, and so forth. I suppose that one can look at parents as neuroscientists with crude, old-fashioned tools! Pereboom continues:

Case 3. Plum is an ordinary human being, except that he was determined by the rigorous training practices of his home and community so that he is often but not exclusively

[22] Pereboom, *Living without Free Will*, pp. 112–113.

[23] Pereboom, *Living without Free Will*, pp. 113–114.

rationally egoistic (exactly as egoistic as in Cases 1 and 2). His training took place at too early an age for him to have had the ability to prevent or alter the practices that determined his character. In his current circumstances, Plum is thereby caused to undertake the moderately-reasons-responsive process and to possess the first-and second-order desires that result in his killing White. He has the general ability to grasp, apply, and regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and hence the rigorous training practices of his upbringing deterministically result in his act of murder. Nevertheless, he does not act because of an irresistible desire.[24]

Finally:

Case 4. Physicalist determinism is true, and Plum is an ordinary human being, generated and raised under normal circumstances, who is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1–3). Plum's killing of White comes about as a result of his undertaking the moderately reasons-responsive process of deliberation, he exhibits the specified organization of first- and second-order desires, and he does not act because of an irresistible desire. He has the general ability to grasp, apply, and regulate his behavior by moral reasons, but in these circumstances the egoistic reasons are very powerful, and together with background circumstances they deterministically result in his act of murder.[25]

Pereboom basically asks the compatibilist to point to the place (after Case 1) along the slippery slope where responsibility emerges. My answer: there is no such place, as Pereboom suggests. Rather, on a plausible understanding of the case, Professor Plum is morally responsible in Case 1. Thus, there is no impediment to saying that Plum is responsible in Case 4 (and, in general, in the context of causal determinism).

As Pereboom points out, Ravizza and I expressed the concern that in certain cases of significant manipulation that occurs literally from birth (or, in this case, from the very beginning of the existence of Professor Plum), there is no opportunity for a self to develop.[26] But let us allow this point to pass, and I shall concede (for the sake of this discussion) that Professor Plum is a genuine self even in Case 1, although created and directly manipulated by others from the beginning. As Pereboom points out, on my view it turns out that Plum has taken responsibility for the manipulation-mechanism; after all, this is the mechanism on which he always acts, and when an individual develops into a morally responsible agent, he takes responsibility for his actual-sequence mechanisms, even if he does not know their details. Further, Pereboom is at pains to point out that the desires on which Plum acts are not irresistible; I take it that Pereboom

[24] Pereboom, *Living without Free Will*, pp. 114.

[25] Pereboom, *Living without Free Will*, pp. 115.

[26] Pereboom discusses this point in the context of a discussion of whether the added dimension of mechanism-ownership can help the Fischer–Ravizza account handle the cases presented above in the text: Pereboom, *Living without Free Will*, pp. 120–123.

wants to say that there is no psychological (or other) *compulsion* here, but mere causal determination. It follows that Plum acts from his own, moderately reasons-responsive mechanism; holding fixed the actual kind of mechanism, there is a suitable range of possible scenarios in which Plum recognizes reasons to do otherwise and does indeed behave in accordance with those reasons.

In this case there is direct manipulation of the brain, but it does not issue in desires so strong as to count as compulsions. Thus, Professor Plum's actual-sequence mechanism has the general power or capacity to respond differently to the very reasons that actually obtain in the case.[27] Although Plum is manipulated by others (without his knowledge or consent), he is not forced or compelled to act as he does; thus, he is not a robot – he has a certain minimal measure of control, and moral responsibility is associated with control (of precisely this sort).[28]

It is crucial here to keep in mind the distinction between moral responsibility and (say) moral blameworthiness (or praiseworthiness).[29] Moral responsibility, as Ravizza and I understand the notion, is more abstract than praiseworthiness or blameworthiness: moral responsibility is, as it were, the "gateway" to moral praiseworthiness, blameworthiness, resentment, indignation, respect, gratitude, and so forth.[30] Someone who is morally

---

[27] For a discussion, see Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 62–91.

[28] In Fischer, *The Metaphysics of Free Will: An Essay on Control*, I made a similar point in regard to God's "providential activity": "Even if God causes human action via a process analogous to causal determination, simply *qua* causal determination (and not *special* causation), then arguably the process can be [suitably reasons-responsive, and the agent morally responsible]" (p. 181).

[29] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 5–8.

[30] The notion of "taking responsibility," a key ingredient of moral responsibility, may (quite understandably) get a "bum rap" from what I might call the "politician's use" of the phrase, "I take responsibility for . . ." Politicians seem to use this phrase precisely as a way of *escaping* accountability or blameworthiness. It is really quite galling. To illustrate the point, consider this amusing story I recently heard told by a comedian (although one can all too easily imagine its being entirely true). A conversation between Jesse Jackson and Bill Clinton takes place after the revelation of Jesse Jackson's marital infidelity. Bill Clinton says, "Jesse, remember what you told me after the public revelation of my infidelity in the Monica Lewinsky fiasco. Recall that you told me that the best way to avoid blame is to take responsibility!"

As Ravizza and I were at pains to emphasize in Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, taking responsibility (on our view) is not merely a matter of mouthing certain words; it is a matter of genuinely having the attitudes in question. One cannot easily avoid blameworthiness by failing to take responsibility. Thus moral responsibility is the gateway to blameworthiness, not a back-door escape.

responsible is an *apt candidate* for moral judgments and ascriptions of moral properties; similarly, a morally responsible agent is an *apt target* for such attitudes as resentment, indignation, respect, gratitude, and so forth. Someone becomes an apt candidate or target – someone is "in the ballpark" for such ascriptions and attitudes – in virtue of exercising a distinctive kind of control ("guidance control"). But it does not follow from someone's being an apt target or candidate for moral ascriptions and attitudes that any such ascription or attitude is justifiable in any given context. After all, an agent may be morally responsible for morally neutral behavior. Further, an agent can be morally responsible, but circumstances may be such as to render praise or blame unjustifiable.

Once the distinction between moral responsibility and (say) blameworthiness is made, it is natural to suppose that Professor Plum is morally responsible for killing Mrs. White, even if he is not blameworthy (or not fully blameworthy) for doing so. After all, Plum is *not* a mere robot – he is *not* compelled or forced to act the way he does. He *does* exercise control, minimal as it may be. It is important to capture this notion of moral responsibility and the associated notion of control, in part because it is important to *mark the difference* between a genuine agent such as Plum (who exercises at least a minimal degree of control) and a robot or individual acting on literally irresistible urges – compulsions. This is the notion of moral responsibility that Ravizza and I aimed to capture.

But it is of course also very important to mark the difference between being morally responsible (in virtue of exercising guidance control) and actually being blameworthy (or praiseworthy). 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 are 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 (or may be, for all that has been said in Pereboom's descriptions) wide disparities in the conditions for blameworthiness.

The ingredients for providing an adequate response to Pereboom's challenge involve the distinction between moral responsibility and (say)

blameworthiness, and the distinction between mere causal determination and action from a compulsive or irresistible urge. One might wonder how to characterize the latter distinction, or whether it exists at all, since (arguably) no desire on which an agent acts can be resisted in a causally deterministic world. I might try to explain the difference, in a rough and ready way, as follows. An irresistible urge is one whose intensity or intrinsic motivational force (whether experienced or not) explains why the action takes place; there is no possible scenario (including those whose pasts differ in their details from the actual past) in which the agent fails to act on the desire, given its intrinsic motivation force. On the other hand, when an agent actually acts on a desire in a causally deterministic world, he may fail to act from a desire with a similar intrinsic motivational force, given differences in the past (or even the laws).

## V. BLACK AND TWEEDALE

To further illustrate this important distinction, let us consider an argument of Sam Black and Jon Tweedale.[31] Black and Tweedale suggest that certain information that we could conceivably receive would make us believe that causal determinism obtains and *thereby* expunge our intuitive sense of our moral responsibility:

> Start by identifying a decision from your past of which you are especially proud or alternatively, especially ashamed. For purposes of illustration, suppose you are an alcoholic and have been a pretty tough nut in all of your fractured personal relationships. Next imagine that you receive a letter informing you that an identical twin separated from you at birth is on their way over to make your acquaintance. As the evening's conversation turns intimate you can't resist asking your twin whether he too has succumbed to those vices for which you are most ashamed (it does not matter whether we focus on your accomplishments instead). You discover that your identical sibling has indeed surrendered to identical vices.[32]

Black and Tweedale contend that you might have mixed feelings about such a discovery. On the one hand, you may feel that you may begin to view your "vices" as no different from "warts or boils – although infinitely more shameful."[33] On the other hand, you might still hold onto the view

[31] Sam Black and Jon Tweedale, "Responsibility and Alternative Possibilities: The Use and Abuse of Examples," *The Journal of Ethics* 6 (2002), pp. 292–306.

[32] Black and Tweedale, "Responsibility and Alternative Possibilities: The Use and Abuse of Examples," p. 294.

[33] Black and Tweedale, "Responsibility and Alternative Possibilities: The Use and Abuse of Examples," p. 294. It is not clear why exactly "shame" would be appropriate, although perhaps the authors are thinking of a shame that does not involve moral responsibility.

that you are morally responsible. Importantly, Black and Tweedale argue, "The second reaction to the example depends for its survival, we suspect, on the tacit assumption that although you and your sibling possess identical vices, your conditions are not causally determined."[34] They elaborate:

> As your conversation progresses into the night even more idiosyncratic shared vices come to light. (These we leave to the reader's imagination.) Once these have been catalogued there comes an insistent knocking, and two (the number is not important) additional identical siblings – reared in similarly independent circumstances – appear at the door. Picking up on the conversation's theme, they too confess to having identical vices. There are now four of you who have made identical messes of your lives – with the possibility of more on the way.[35]

They continue:

> ... when the peculiarities of our personality are viewed in this light they seem no different from the oddities of our physical appearance, such as our height, hair or eye color; that is to say, as natural facts about us for which we take neither credit nor blame. ... If these reflections are on the right track they support incompatibilism. For the incompatibilist claims that discovering the existence of an identical twin is like discovering the causal determinants of our behavior. The appearances of successive siblings simply render the causal determinants of our behavior increasingly transparent. But in principle we should reach the same conclusion about moral responsibility any time we fully appreciate how the course of someone's deliberations is uniquely determined.[36]

Now it seems to me that this sort of evidence would be in favor of the conclusion that our behavior generally (or always) issues from irresistible desires. What would make such evidence so surprising – indeed, startling – is that it would point to the conclusion that all our behavior is the result of irresistible urges or *compulsions*. Such evidence would *not* be evidence for *mere* causal determination of behavior; it would be evidence that our genes somehow *compel* us to act, even if we are unaware of such compulsion). *This* is why we would find such hypothetical and wildly implausible evidence so startling. It is not the *mere* thought that our choices and behavior is causally determined that is shocking, but rather the thought that all our choices and behavior are *compelled*. At the very least, thought experiments involving hypothetical evidence about identical twins cannot *in itself* show that we would be startled to find that our behavior is causally determined (and that we would thus give up our view of ourselves as morally responsible persons).

---

[34] Black and Tweedale, "Responsibility and Alternative Possibilities: The Use and Abuse of Examples," p. 294.

[35] Black and Tweedale, "Responsibility and Alternative Possibilities: The Use and Abuse of Examples," pp. 294–295.

[36] Black and Tweedale, "Responsibility and Alternative Possibilities: The Use and Abuse of Examples," p. 295.

Return, now, to Pereboom's Professor Plum of Case 1, who we discussed above. Let us suppose that as a young man, as he was developing into a morally responsible agent, he took responsibility for his "ordinary" mechanism of practical reasoning (which involves the covert manipulation by the neuroscientists). Many years later (say three decades), he acts from this mechanism, which is, by hypothesis, moderately reasons-responsive. As I said above, I am inclined to say that Plum is morally responsible for killing Mrs. White, although most likely not blameworthy (or significantly blameworthy). I would distinguish Plum from Professor Glum, who is *not* manipulated as a young man, and takes responsibility (when a young man) for the exercise of the ordinary human capacity for practical reasoning. Later in his life (say three decades later) the neuroscientists begin to manipulate him in a clandestine fashion. A week later, he acts on this mechanism (that involves covert, undetected manipulation by the scientists) in just the same fashion as Plum: he kills Mrs. White, and the operations of his brain and body are isomorphic to those of Plum. We can even assume that Glum's configuration of character traits and motivational states are such that it is plausible to suppose that he would have killed Mrs. White in just the same way in which he actually kills her, if he had not been manipulated by the neuroscientists. I believe that, whereas Plum is morally responsible for killing Ms. White, Glum is not. Plum acts from his own, moderately reasons-responsive mechanism, but Glum does not. Glum's actual-sequence mechanism is not his own – he has not taken responsibility for the manipulation-mechanism.

I concede that it may not be obvious that my intuitions about these cases are correct. Perhaps it will be thought that my theory is driving my intuitions here, rather than the other way round. I do not know how to establish that my intuition is correct, or that it is largely independent of my theory. I can simply display the results of my theory in these cases, and profess my agreement. What may, however, be helpful is that the asymmetry between Plum and Glum (on my approach) shows that the Fischer–Ravizza theory of moral responsibility is "historical" in a strong way.

## VI. ZIMMERMAN

To explain. Some years ago my co-author and I suggested that the notion of moral responsibility is (like justice, love, and other notions) an *essentially historical* notion.[37] We contrasted historical notions with those that are

[37] Fischer, *The Metaphysics of Free Will: An Essay on Control*, Fischer and Ravizza, "Responsibility and History," and *Responsibility and Control: A Theory of Moral Responsibility*, pp. 170–206.

"current-time slice" notions, such as shape, color, weight, and so forth. You can tell an object's color by looking at it and noticing its current time-slice characteristics. You cannot tell whether an agent is morally responsible by simply considering the agent's current time-slice properties, such as his configuration of mental states. Various philosophers have pointed out that this dilemma is not exhaustive; there can be "process notions" that are neither current time-slice nor deeply historical notions.[38] Perhaps it takes awhile to "identify" with a particular first-order desire; perhaps, for example, this process of identification involves (at least) the formation of a higher-order desire to act in accordance with that first-order desire. Roughly this sort of account, suitably filled in and elaborated, is not exactly a current time-slice model; nor is it historical in a particularly interesting or deep way.[39] One simply has to focus on a suitable *interval*, rather than an instantaneous time-slice.

This is a good and helpful point. Of course, such "process-accounts" remain problematic, because manipulation can occur *over the relevant interval*. So, although they are not, strictly speaking, current-time-slice models of moral responsibility, they are equally open to the manipulation objection. More to my purpose here, it should be evident from the asymmetric treatment of Professors Plum and Glum that my account of moral responsibility is not merely a process-notion, but it is historical in a deeper way. Plum and Glum choose and act in exactly the same way; on the Fischer–Ravizza account of moral responsibility, the difference in their responsibility-status comes entirely from events that occurred decades earlier – events that are not plausibly thought to be parts of an extended responsibility-conferring process. Additionally, those events (the taking-responsibility events) are not themselves exercises of guidance control that are related to future behavior in the way that (say) freely getting drunk is

[38] For probing discussions of this set of issues, see Gary Watson, "Some Worries about Semi-Compatibilism: Remarks on John Fischer's *The Metaphysics of Free Will*," *Journal of Social Philosophy* 29 (1998), pp. 153–143, and "Reasons and Responsibility," *Ethics*, 111 (2001), pp. 383–386; and David Zimmerman, "Reasons-Responsiveness and Ownership-of-Agency: Fischer and Ravizza's Historicist Theory of Responsibility," *The Journal of Ethics* 6 (2002), pp. 199–234, and "That Was Then This Is Now: Personal History vs. Psychological Structure in Compatibilist Theories of Autonomous Agency," *Nous* (forthcoming).

[39] This sort of hierarchical account was suggested (in contemporary philosophy) by Harry Frankfurt in "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68 (1971), pp. 5–20; it has subsequently been developed in additional essays by Frankfurt, and discussed widely.

related to future out-of-control driving. My theory of moral responsibility, then, is genuinely and deeply historical.[40]

Moral responsibility is in this respect like love. The notion of love is quite mysterious, as is love itself.[41] In understanding the notion of love, and its distinctive "particularity," it is helpful to begin with two features: its essential historicity and non-fungibility (I will add a third dimension below). The historicity of love entails that there cannot be love at first sight. A certain sort of history must be shared, in order to have genuine love. Thus, there cannot be literal "love potions," just as there cannot be "virtue pills." The non-fungibility of love entails that if one loves a beloved, and the particular beloved changes (i.e., the object of the attitudes constitutive of love is a different particular person), then one does not any more have *love* toward that new individual. This is of course compatible with there being changes, even radical changes, in the *properties* of the beloved (consistent with the continuation of love).

Imagine that your spouse (I will say, "wife") and three children are all hit by lightning bolts as you are driving home from work. By some inexplicable cosmic accident, there emerge molecule-for-molecule doppelgangers of them – with all of the same properties (mental states, dispositions, memories, and so forth) of the originals. The new individuals – and they are new, for there is no connection at all between the original persons

[40] David Zimmerman suggests that in order to have a plausible, deeply historicist approach to moral responsibility one must address a certain fundamental question: "How do some children manage to develop the capacity to *make up their own minds* about what values to embrace, by virtue of having gone through a process in which they play an increasingly active role in *making their own minds*, a process which begins with their virtually *having no minds at all*?" (Zimmerman, "Reasons-Responsiveness and Ownership-of-Agency: Fischer and Ravizza's Historicist Theory of Responsibility," p. 233) Addressing this question would be perhaps crucial as part of an overall theory that encompassed both moral responsibility and also an account of the conditions of blameworthiness and praiseworthiness; but our goal in presenting the account of moral responsibility was not so lofty. In order to provide a complete theory that includes a specification of the conditions of blameworthiness, praiseworthiness, indignation, resentment, resentment, and so forth, one would need to have an account of autonomous value and preference-formation; but we did not set out to give such an account. An account of the kind of control required for moral responsibility need not address the very fundamental, and dauntingly difficult, question of the different between (say) indoctrination and education, or, at the very basic level, autonomous value formation. Whew!

[41] In recent work, Harry Frankfurt has given a particularly perspicuous and nuanced account of love: Harry Frankfurt, *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999).

and the replacements – await you at home. If you knew what has happened, what should your reaction be, and how should this be characterized?[42]

Ravizza and I have argued that, since love is essentially historical, it would be inappropriate to characterize your attitudes to the new individuals as love (at first). A period of time during which you interact with the new individuals is necessary. This also follows from the non-fungibility of love. But it seemed to us that it would be unbearably harsh and cold to suppose that you should not have attitudes and feelings toward the new inidivudals not unlike those toward the originals. After a suitable period, these attitudes could properly be described as love (rather than something like "proto-love,"), and one can properly be said to love the new individuals.

David Zimmerman criticizes the above treatment of the notion of love. He believes that it indicates an inappropriate understanding of the deeply historical nature of love (parallel to our alleged misunderstanding of the deeply historical nature of moral responsibility):

> I doubt that Fischer and Ravizza's . . . position is plausible (even if coherent), for the essential historicity of *adult love at time t* seems (to me, anyway) inextricable from the fact that the lover has shared a history with *this particular non-fungible* beloved. To be sure, there is room (just barely) in our lives for a relational emotion which involves a shared history only with a bundle of properties *however instantiated* by particular persons at various stages of the particular lover's history. Call this "Love *de dicto*." A vivid example would be the James Stewart character's obsessive efforts in "Vertigo" to "recreate" his "Madeleine" (the first Kim Novak character). But a lover who is *aware* of the replacement of the original instantiating particular person and who continues to have all the same old feelings toward the new instantiation of the same set of type-identical properties as he did toward the original, like the husband for his "replacement wife" in Fischer and Ravizza's doppelganger example, is surely suffering from a kind of pathology beyond mere fickleness.
>
> . . . [Fischer and Ravizza's position] brings out yet again the importance of distinguishing between the mere *process* and the deep *source* dimensions of conceptually historical properties. For the reply makes it sound as though the enduring instantiation of the former beloved properties, *never mind how*, is what does the trick. But surely if contemporary interaction can transform mere proto-love into the genuine article, it does so not simply by virtue of the lover's becoming accustomed to the idea that the beloved set of properties is instantiated anew in a doppelganger replacement, but rather by virtue of the fact that he shares enough time with *this particular* "proto-beloved" so that this very interaction can be the source of *new* lovable properties in both of them.
>
> Amelie Rorty suggests that "love is not love that alters *not* when it alteration finds" because the genuine article has to be open to the possibility that the loves will so change as a result of dynamic interactions which occur during their shared history (both between

[42] This thought-experiment comes from Mark Bernstein, "Love, Particularity, and Selfhood," *The Southern Journal of Philosophy* 23 (1985), pp. 287–293. It is discussed in Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 192-94. Originally, the suggestion that love is historical was made by Robert Nozick: Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 67–68.

them and between each lover and the rest of the world) that one or the other might fall out of love. I offer a (less poetic) corollary: "love is not love for would-be lovers who in the fullness of time do not alter when they replacement find." But this is a *source*-historicist condition, for it requires not only that the husband who is made newly aware of the replacement be afforded some time to get used to the idea that this instantiation of the beloved properties now interacting with him is a doppelganger, but also that the new phase of his historical interaction with his proto-beloved replacement be a potential source of at least some new relational properties of both of them. In other words, emotional *stasis* after the husband becomes aware of the replacement entails that he does not really love the doppelganger wife but just a bundle of properties, however instantiated.[43]

But there is absolutely nothing in the Fischer–Ravizza approach to the puzzle about replacements that entails (or, as far as I can see, even suggests) the sort of "emotional stasis" described by Zimmerman. On our view, you should still have the sorts of general attitudes characteristic of love toward the new individuals; the attitudes simply cannot be described (yet) as love (or part of love). Love is historical, and its object is non-fungible.

In the replacement case, as you interact (say) with your replacement spouse and have many of the general attitudes characteristic of love, the relationship may mature and develop into genuine love. Of course, as with love of one's original spouse, this may include an openness to changes in the interests and personality of the spouse. Nothing in our view precludes this sort of openness, and an associated appreciation for change and development in your beloved.

I have tried to defend a certain view of love as historical in a deep sense. This is not unlike the Fischer–Ravizza view of moral responsibility as deeply historical. I have suggested that the historicist nature of love is a component of the more general *particularity* of love. Love's particularity consists at least in its essential historicity and the non-fungibility of its object. I want finally to suggest that there is a third dimension, perhaps difficult to articulate, of love's particularity; this dimension pertains to its individuation, as it were. Having interacted suitably with the replacement spouse, one can actually be said to love her. But this is not the *same love* – it is a different love because it has a different object.

One can speak of "the great loves of one's life." It may be that one is simply pointing to different beloveds. Or it may be that one is indicating different *instances* of love (where the "instances" are not instantaneous, but take place over durations). Love is particular in the sense that it is defined in terms of general attitudes and also a particular beloved; when the

[43] Zimmerman, "Reasons-Responsiveness and Ownership-of-Agency: Fischer and Ravizza's Historicist Theory of Responsibility," pp. 231–232.

particular beloved changes, even apart from any changes in general properties (interests, character traits, and so forth), there is a different *instance* of love. In the replacment puzzle, your love for your family constitutes a *regulative ideal*: it impels you to have the same general attitudes, including an appreciation of and openness to change in the individuals who are the targets of the attitudes, and it ultimately points to new love.[44]

I began the discussion of love by remarking on its mysteries. The ruminations above remind me of that great, old country and western song, "I Don't Know Why I Love You, But I Do."

## VII. MECHANISM-INDIVIDUATION: MCKENNA

The overall theory of moral responsibility that Mark Ravizza and I presented has various components: the contention that moral responsibility does not require the sort of control (regulative control) that involves metaphysically open alternative possibilities, the claim that guidance control is the freedom-relevant condition necessary and sufficient for moral responsibility, the idea that guidance control can be analyzed in terms of mechanism-ownership and moderate reasons-responsiveness, and the claim that guidance control, so construed, is compatible with causal determinism. Of course, these elements can be further broken down into their parts; for example, moderate reasons-responsiveness is analyzed in terms of "sameness of mechanism," regular reasons-receptivity, and weak reasons-reactivity.[45] A part of the overall theory that we conceded to be vague, and which has been fixed on by various commentators, is the notion of "sameness of kind of mechanism."[46]

---

[44] There is a helpful and penetrating alternative account of love's particularity in Robert Adams, *Finite and Infinite Goods* (Oxford: Oxford University Press, 1999), pp. 131–176. If I may explicate Adams' view in an over-simple way, I believe that Adams holds that one loves another particular individual by first loving certain tropes – certain property instances (her courage, her sensitivity, and so forth). Loving the tropes is prior, and one constructs love of general properties from love of the tropes. In this way love is particular.

[45] For the latter notions, see Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 62–91.

[46] For particularly forceful and penetrating discussions, see: Michael McKenna, "Review of John Martin Fischer and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*," *The Journal of Philosophy* 98 (2001), pp. 93–100; and Gary Watson, "Reasons and Responsibility: Review Essay on John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*," *Ethics* 111 (2001), pp. 374–394.

The theory, as presented by Ravizza and me, does not contain an explicit account of mechanism-individuation. We acknowledged this fact, and conceded that it is a potential problem.[47] I want to say a bit more here about the role that this fact plays in the theory – and the assessment of the theory. I shall begin by laying out the critique developed by Michael McKenna. In doing so, I want to address (at least in a preliminary way) McKenna's challenge:

> Fischer and Ravizza's appeal to sameness of mechanism is the lynchpin in their defense of an actual-sequence, reasons-responsive analysis of guidance control. Regrettably, their exclusive reliance on intuition as a basis for mechanism individuation renders their defense of their overall theory unconvincing. There are too many pressure points at which differing intuitions regarding sameness of mechanism yield troubling results for their defense of guidance control. Thus, to defend their compatibilist account of moral responsibility fully they must address this source of trouble.[48]

McKenna elaborates the worry as follows:

> ... because they [Fischer and Ravizza] offer no *principled* basis for mechanism individuation, they must rest their thesis purely on *intuitive* reactions to different cases. But, it might be objected, *which* elements from the entire complex (of proximal events and states antecedent to an action) should figure intutitively into the relevant mechanism will vary relative to explanatory perspective. The neurophysiologist's basis of parsimony will be different than that employed in everyday folk-psychological discourse. What reason have we to assume that Fischer and Ravizza's basis for individuation is the correct one?
>
> The situation worsens if one pushes for a hyper-resptricted notion of sameness of mechanism. On the hyper-restricted construal, the entire complex of proximal antecedent events and states function as the pertinent mechanism. If this were the relevant mechanism, an agent could not act from a reasons-responsive mechanism at a deterministic world.[49]

I agree that a *full defense* of our compatibilistic approach might well involve a "principled" account of mechanism individuation. Without such a defense, I fully concede that the overall theory, and its "defense," is *incomplete* (I prefer that word to "unconvincing"). But I also would suggest that it is unreasonable to expect that anyone could present a defense of a highly contentious thesis about free will, *all* of whose elements are decisively and uncontroversially defended (via appeal to "principles" rather than intuitions). I am not sure exactly how one could produce a purely "principled" account of mechanism individuation – an

[47] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, p. 40.

[48] McKenna, "Review of John Martin Fischer and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*," p. 100.

[49] McKenna, "Review of John Martin Fischer and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*," p. 97.

account that did not at some level appeal to intuitions. It is obvious that the notion of "mechanism leading to action" is quite vague in itself, and open to various interpretations that depend on various "explanatory perspectives." And, in general, I think that interesting attempts at solving genuinely difficult philosophical questions will often be incomplete and dependent to some extent on intutions, rather than general principles.

Surely it would be setting the bar too high to demand that any candidate for a solution to a philosophical puzzle must have all of its components defended in a fully general way, with no vagueness, no fuzzy edges, and no appeal to intuitions. I am afraid that this would limit the candidates rather drastically! On the other hand, it is quite fair and legitimate to point out that there is an important incompleteness in the theory of moral responsibility sketched by Ravizza and me, and to press the issue of whether the vaguenessss of the notion of "sameness of kind of mechanism" allows the proponent of the theory to allow his intuitions, rather than the theory, to do all (or most) of the work. That is, it is a perfectly reasonable worry that we simply apply the theory in such a way to get the results that match our intuitions, exploiting the vagueness of "sameness of kind of mechanism" to come down one way in this case, another way in that one, and so forth.[50] If this were so, then the theory really would not be illuminating and systematizing our intuitions – it would simply be a front for them.

This worry raises deep and difficult methodological and substantive questions. I can only gesture at a response, in the most preliminary of ways. First, the structure of our theory of moral responsibility – in which one holds fixed the "actual-sequence mechanism" – is similar to the structure of "reliabilist" theories of knowledge.[51] In these theories, ascertaining whether an individual has knowledge involves holding fixed the actual-sequence belief-producing mechanism and asking whether it is "reliable" – whether, for instance, it tracks truth (in Robert Nozick's terms).[52] Indeed, since Nozick offers no general account of mechanism-individuation (of belief-producing mechanisms), he is aware of a problem

---

[50] For interesting and subtle cases that press essentially this concern, see Watson, "Reasons and Responsibility: Review Essay on John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsiblility*," pp. 379–383.

[51] I discuss certain aspects of this isomorphism in Fischer, *The Metaphysics of Free Will: An Essay on Control*.

[52] Robert Nozick develops this sort of theory of knowledge, and points out the structural isomorphism with a theory of "tracking bestness" (which is not exactly an account of moral responsibility), in Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981), pp. 167–362.

for his theory of knowledge which is parallel to the problem about mechanism-individuation I described above.[53]

Just as Nozick is not convinced that he is guilty of putting the cart before the horse, as it were, I am not convinced that the vagueness of our notion of mechanism-individuation renders our theory of moral responsibility otiose. Various philosophers have offered penetrating and challenging criticisms of "reliabilist" accounts of knowledge, which press concerns about mechanism-individuation. I do not know whether these critiques are decisive; I certainly think that reliabilist approaches in epistemology are illuminating and worthy of serious consideration, even if one wants to reject them ultimately (because of the worries about mechanism-individuation, or for other reasons). Further, I have not seen *any* argument that contends that our actual application of our theory of moral responsibility to cases is problematic in the ways in which reliabilism in epistemology is (allegedly) problematic.

Any theory which involves *generality* appears to have problems, at some level, of the sort we have been considering. Rule-consequentialism (of which rule-utilitarianism is an example) and Kantianism (in ethics) are salient examples (along with reliabilism in epistemology) of theories that are "generalizing" theories. Rule-consequentialism asks what the consequences of a general acceptance of a certain rule would be, where the rule specifies *kinds* of acts. Kantianism asks whether it would be (say, logically) consistent for all agents to act in certain ways – motivated by certain *kinds* of maxims or intentions. Typically (although perhaps not universally), reliabilists, rule-consequentialists, and Kantians do not offer reductive, general accounts of the individuation of the relevant "kinds." At some level they rely on intuitions; they implicitly adopt approaches to individuation that help the theory yield the "right" results. Surely, generalization approaches in ethics, as well as reliabilism in epistemology, are serious, illuminating approaches, which should be taken seriously, even if they are ultimately rejected. I would hope that the theory of moral responsibility in terms of guidance control, as sketched by Ravizza and me, could be similar to the other generalizing theories at least in the respect that it may be considered to be illuminating and worthy of serious consideration. I would hope that it could be seen to throw into relief a whole host of traditional issues, restructuring some of the traditional debates in a way that makes them more tractable, or, at a minimum, makes the precise points of disagreement more perspicuous.

Finally, I want to emphasize a feature of the methodology employed by Ravizza and me that helps to provide an answer to the worries pressed

[53] Nozick, *Philosophical Explanations*, pp. 179–185.

by Watson and McKenna (and others) about mechanism-individuation. I am afraid that we did not highlight this sufficiently, and our defect in this regard has led to some unclarity about our goals. I hope to help to clarify our position here. In *Responsibility and Control*, we write:

> ... we aim to give what Robert Nozick has called 'philosophical explanations,' not to do 'coercive philosophy.' That is, we will be seeking to show that it is very plausible and appealing to say that (for example) agents can be held morally responsible for their behavior, regardless of the truth (or falsity) of causal determinism. And we will be trying to show exactly *how* this sort of view can be developed and defended. But we do not suppose that we can give a knockdown argument for this conclusion (or the other major contentions of the book). Thus, when we contend that we have argued successfully for (say) the compatibility of causal determinism with moral responsibility, we are claiming that we have offered a strong plausibility argument for this conclusions, but not an argument that any rational agent is compelled to accept.[54]

We go on to point out that we are seeking to systematize our society's shared consensus about cases in which certain factors undermine moral responsibility – and to distinguish them from cases in which no such uncontroversial responsibility-undermining factors operate.[55]

So the overall dialectical structure of our argument can be limned as follows. We offer what we take to be strong plausibility-arguments for the claims that moral responsibility does not require alternative possibilities, and that causal determinism in itself does not rule out moral responsibility.[56] We then offer a general theory of moral responsibility that shows how it is *possible* to defend, in detail, these views – in particular, that moral responsibility is compatible with causal determinism. This theory gains some credibility from its *systematic and unified* treatment of moral responsibility for actions, omissions, consequences, and even traits of character. Of course, our arguments for the overall approach are not *decisive*, and various elements remain to some degree or another vague and undeveloped. The vagueness in the notion of mechanism-individuation allows us to apply the account of guidance-control in such a way as to match our considered judgments about the cases. In a sense, we here allow our intuitions to guide us in that they point to the way of individuating mechanisms, if our theory is to "work." This is part of the project of showing in some detail how it is *possible* to defend a kind of compati-

[54] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, p. 11.

[55] Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, pp. 34ff.

[56] These arguments are offered in our work as a whole, including Fischer, *The Metaphysics of Free Will: An Essay on Free Will*, as well as Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.

bilism about causal determinism and moral responsibility, and, as far as I can see, it does not imply any sort of problematic inconsistency or circularity.

Of course, it follows that we cannot convince a committed incompatibilist of the truth of compatibilism (by invoking the theory, as developed thus far). But this is no big surprise. We never supposed that we could *prove* compatibilism – we did not set out to do coercive philosophy. It is a big enough job, I think, to show exactly why it would be desirable if compatibilism turned out to be true, why compatibilism (about causal determinism and moral responsibility) does not involve obviously unacceptable commitments (in contrast, perhaps at least, to compatibilism about causal determinism and freedom to do otherwise), and how – in some detail – one might present a systematic compatibilist theory.

## VIII. WATSON'S CHALLENGE AND THE DIFFERENT MODALITIES

Gary Watson has posed a particularly pointed challenge – one that goes to the very heart of our theory of moral responsibility:

> It is also somewhat curious that Fischer and Ravizza feel the need to make this modal claim [the claim that, when an agent is morally responsible, the mechanism on which he acts has the general capacity to respond to the actual reasons]. The objection regarding fairness seems to arise from intuitions supporting a principle of alternative possibilities (holding people responsible is unfair unless they could have done otherwise). Fischer and Ravizza reject this principle because of so-called Frankfurt cases, in which some fail-safe device stands by to ensure that an individual behaves in a certain way. For example, suppose that if Goldie were to change her mind at the last moment about voting for the Green candidate, the fail-safe device would ensure that she punched the "Nader" tab anyway. So, there is no possibility that she would not punch that tab. Fischer and Ravizza reasonably conclude the this modal fact does not entail that her actual voting behavior is not reasons-responsive. This leads them to reject the idea that to be responsible, the agent must have alternatives to what she does. In Frankfurt cases, Fischer and Ravizza like to say, the agents could not have responded differently in the face of contrary incentives, but the actually opeative mechanisms could have ...
>
> What is curious, then, is that Fischer and Ravizza seem to feel the need to employ a notion of alternative possibilities at the level of mechanisms. They seem to be conceding that there is a sense in which the fairness of holding someone responsible depends upon the capacity of the mechanism in question to respond otherwise, a capacity that must be compatible with causal determinism, on their view. But it is hard to see how this move can answer the concern about fairness, unless we can translate talk about the capacities of mechanisms into talk about what persons can do. And if we can do that, we should endorse a compatibilistic version of the principle of alternative possibilities rather than rejecting the relevance of alternative possibilities altogether.[57]

[57] Watson, "Reasons and Responsibility: Review Essay on John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*," p. 382.

This is a probing and difficult challenge. In seeking to respond, I begin by noting an analogy between the active power, freedom, and certain passive powers, such as (say) solubility in water. As I have pointed out in previous work, Frankfurt-type examples are just one kind of example of "Schizophrenic Situations." Objects in Schizophrenic Situations can exhibit either active or passive powers – these situations contain a kind of "swerve" in metaphysical space. One can construct the analogues of Frankfurt-type cases for passive powers.[58]

Consider, for example, Alvin Goldman's example of a piece of salt, which is an ordinary piece of salt, with an ordinary structure (in virtue of which it is soluble in water); what is unusual is that there is a magician associated with this piece of salt, and if the piece of salt were about to be placed in water, the magician would waive his magic wand and cause the salt to have an impermeable coating. So the salt actually displays a structure in virtue of which it is plausibly thought to be soluble in water; but it is not the case that it would dissolve, were it placed in water. Given the presence of the magician, and the fact (let us suppose that it is a fact) that the magician cannot be distracted or otherwise deterred, this particular piece of salt cannot dissolve in water. And yet it seems to be water soluble. It is water soluble in virtue of actually displaying a certain sort of structure – a structure that underwrites a general capacity.

An approach to analyzing the water solubility of such a piece of salt would be to hold fixed the actual structure of the piece of salt (i.e., the structure *sans* special impermeable coating), and to ask what would happen if the salt is put into contact with water (given that the magician does not intervene). This is an actual-sequence approach to analyzing the passive power, solubility, which is parallel to the analysis of the active power, guidance control. In both cases the general capacity which is actually displayed or exhibited is held fixed under counterfactual circumstances (in which other factors are allowed to vary). I suppose one could object that this is an untenable or analytically unstable analysis of water solubility. One could say that the piece of salt is not really soluble in water, since it cannot dissolve in water: it would not dissolve, if it were placed in water. Why focus on the general capacity of salt with the actually-displayed structure, if *this* piece would not display that structure, if it were placed in water? And if we choose to say that this piece of salt is indeed water-soluble in virtue of actually displaying a certain structure (and thus general

[58] I discuss Schizophrenic Situations, and the associated swerve in metaphysical (or logical) space, in Fischer, *The Metaphysics of Free Will: An Essay on Free Will*, pp. 154–158. Alvin Goldman presented his piece of salt example in, Alvin Goldman, *A Theory of Human Action* (Englewood Cliffs: Prentice-Hall, 1970), pp. 199–200.

capacity), why not define a notion of "possibility" relative to which this piece of salt *can* dissolve in water?

I do not know how to *argue* for the contention that a piece of salt that actually has the normal chemical structure of salt is water-soluble, even if it has a weird magician of the sort described above associated with it. I do think that anything actually having the normal chemical structure of salt is soluble in water. I do not think that there is anything analytically unstable about defining water-solubility in terms of this actually-displayed structure (and general capacity), while noting that the particular piece of salt cannot dissolve in water. I suppose that one might define a notion of possibility that abstracts away from "obstacles" or conditions that would prevent the manifestation of a certain dispositional property, and then employ this notion of possibility to say that, yes, Goldman's piece of salt can indeed dissolve in water. Whereas I do not see exactly what is gained by this move, it is certainly available.

I have invoked the analogy between active and passive powers to suggest that at a certain level of analysis there is nothing problematic or unstable about fixing on the general capacity that is actually exhibited, while noting that the object in question lacks a certain sort of power to do (or be) otherwise. This sort of analysis is, I believe, natural and plausible for passive powers, and I would suggest that it is similarly attractive for active powers (such as freedom or guidance control).

But Watson's challenge pertains more specifically to "fairness." How is it fair to hold an agent morally responsible for acting on a general capacity that is indeed sensitive to the particular reasons that actually obtain, even where the agent cannot respond to that reason? I do not know how fully to address this worry, but I would at least sketch the following idea.

Clearly, an individual can act in a way that is not a manifestation of a particular trait of character or general capacity. A courageous person may act in a cowardly manner in a particular situation. In this situation, the cowardly act does not exhibit or display the trait of courage. Whereas the person may be commendable for his courage, we hold him responsible, in the context in which he acts in a cowardly manner, precisely for his cowardly behavior. Similarly, an agent may not act in such a way as to manifest the general capacity for moderate-reasons-responsiveness – he may act from a compulsion or because of direct stimulation of the brain, and so forth. But when an agent does manifest this sort of capacity, he *links* or *connects* himself with this capacity in a distinctive way. In forging this link or connection, the agent is, as it were, inviting (or, in effect, allowing) others to treat him *as acting from this sort of mechanism*. In reacting to the agent's behavior (and thus holding him responsible), we

are thus justified in replying to the agent *qua* agent-acting-from-the actual-sequence mechanism. Thus, considerations of fairness *shift* from the agent to agent-qua-practical-reasoner-of-a-certain-sort. If we are considering the agent-as-acting-from-a-certain-general-capacity, we want to know whether the general capacity that is actually displayed can respond to the actual incentives. (Similarly, when we are considering a piece of salt *qua*-piece-of-salt-with-the-actually-displayed-structure (and thus general capacity), we want to know whether a piece of salt with *that* structure and capacity would indeed dissolve in water.)

On my approach to moral responsibility, I focus on the general capacity for reasons-responsivenss actually displayed by the agent. I contend that an agent can exhibit a suitable sort of reasons-responsivenss (and guidance control), even if the agent could not have done otherwise (and thus does not possess regulative control). But once one makes the move to actually displayed general capacities, why not *also* define a notion of possibility relative to which the agent *can* do otherwise? So we could say that the agent *qua* practical-reasoner-of-a certain-sort could have done otherwise, even in a Frankfurt-type case, just as the piece of salt-*sans*-intervention-by-the-magician could have dissolved in water, in the Goldman-type case.

As I pointed out above, I do not see that anything is gained in terms of analytical penetration by making this sort of move. But I do not have any strong objection to pointing out that the agent-*qua*-practical-reasoner-of-a-certain-sort (i.e., *qua*-acting-on-the-actual-sequence-mechanism) can do otherwise in the Frankfurt-type case. What *would* be objectionable would be to conclude from this that the agent can, in the ordinary sense of "can in the particular circumstances," do otherwise (in the Frankfurt-type case).

There is nothing problematic, as far as I can see, in fixing on the actually displayed general capacity (and its modal characteristics) in the context of causal determinism. That is, there is nothing problematic, in my view, in contending that the relevant agent acts freely (exhibits guidance control) in such a context. In contrast, if one says that the agent could have done otherwise (possessed regulative control), then one must say that the agent could have either so acted that the past would have been different from what it actually was, or the laws would have been different from what they actually are. So there is the following important asymmetry between imputing regulative and guidance control in a causally deterministic context: attributing regulative control requires an answer to the powerful skeptical arguments flowing from the fixity of the past and natural laws, whereas attributing guidance control does *not*.

My theory of moral responsibility has a specific modal structure. I have called it an "actual-sequence" theory of moral responsibility. This means that I do not require that agents have genuine access to alternative possibilities – they need not have regulative control. On the other hand, I *do* require that morally responsible agents act from actual-sequence mechanisms that are *moderately reasons-responsive* – i.e., actual-sequence mechanisms that have certain *modal* or *dispositional* characteristics.[59] Note that this puts my approach – semi-compatibilism – in the mid-point of a certain spectrum. On the one hand, the libertarian argues that moral responsibility requires regulative control – I deny this. On the other end of the spectrum, R. Jay Wallace argues that moral responsibility does not require such control, but simply requires the possession of the general capacity for reasons-sensitivity, not necessarily the actual display of this capacity. My view is in the middle: I argue that moral responsibility requires not just the possession of a certain general capacity for reasons-sensitivity, but the actual display of such a capacity: moral responsibility requires action from a mechanism that is (in addition to being the agent's own) moderately reasons-responsive.

## IX. Concluding Remarks

I (together with my co-author, Mark Ravizza), have sought at least to provide the skeletal structure of an overall approach to moral responsibility. This approach is distinctive in that it is an "actual-sequence" approach; that is, we do not require the sort of control that involves genuine access to alternative possibilities at *any* point: in forming character, performing actions or omitting to act, and bringing about consequences. In developing this overall theory, we fix exclusively on features of the actual pathways to the behavior (or character traits), albeit (sometimes) modal or dispositional features of these pathways. It is an actual-sequence approach in that we do not require alternative possibilities. It may or may not be the case that the future is a garden of forking paths (depending in part on whether or not causal determinism obtains), but this does not matter for moral responsibility.

---

[59] So what happens in other possible worlds is not irrelevant to one's moral responsibility. On my view; rather, what happens in other possible worlds is relevant *not in virtue of pointing to regulative control, but only in virtue of specifying the modal characteristics of the actual sequence mechanisms that potentially count as part of the agent's guidance control.*

The approach is a cohesive package, consisting of various separable parts. The parts themselves contain parts (in some instances). We have offered arguments for some of the parts, but have not been able to offer explicit arguments for all components (or their elements). A basic motivating engine of semi-compatibilism is that moral responsibility, and even personhood (robustly construed), should not depend on whether the formulas that physicists develop (to describe the world) are univeral generalizations or merely almost universal generalizations. The fundamental differences between persons and nonpersons, and morally responsible agents and those individuals who are not morally responsible, should not hinge on arcane deliverances of theoretical physicists – we should not have to stop treating other human beings as deeply different from other animals (and computers) if a consortium of scientists discovers the truth of causal determinism.

Against the background of this motivation, we argue that moral responsibility (and personhood) does not require regulative control. Thus, some of the most disturbing arguments for the incompatibility of causal determinism and moral responsibility are rendered irrelevant. We go on to consider *other* arguments for this sort of incompatibilism, and find none of them compelling (or even strong). Given this dialectical niche, we present an overall, systematic compatibilist account of moral responsibility. On this approach, the freedom-relevant condition necessary and sufficient for moral responsibility is guidance control, and the conditions for responsibility for actions, omissions, consequences, and even traits of character are *tied together* in a unified way.

The account of guidance control assumes a certain intuitive way of individuating the kinds of mechanisms that issue in behavior; we concede that we can offer no entirely "principled" way of individuating mechanisms. In my view, this shows that the overall approach is incomplete, but not fatally flawed. The specific account of guidance control we offer shows how it is possible to develop a compatibilist account of moral responsibility, but it clearly (in itself) does not *justify* or *establish* compatibilism.

Here I have tried to address some of the most penetrating and illuminating criticisms of the overall approach. In doing so, I have sought to clarify the theory. This clarification has in some instances revealed the goals of the theory to be different from, and perhaps less lofty than, those attributed to it by its critics. For example, Ravizza and I seek to give an account of moral responsibility, but not (yet) a full account (say) of praiseworthiness and blameworthiness. Also, we do not aim to *prove* or *establish* compatibilism, but to motivate it and to show how it can be developed in a

coherent, attractive way. Of course, if one's aims are sufficiently modest, this renders the views immune to critical assault – but one purchases this immunity at the cost of not saying anything of interest. I hope that we have found the right mix of humility and daring.

*Department of Philosophy*
*University of California*
*Riverside, CA 92521*
*USA*
*E-mail: John.Fischer@ucr.edu*