JOHN J. DAVENPORT

FISCHER AND RAVIZZA ON MORAL SANITY AND WEAKNESS OF WILL

(Received 13 October 2000; accepted in revised form 7 May 2002)

ABSTRACT. This essay evaluates John Martin Fischer and Mark Ravizza's mature semi-compatibilist account of moral responsibility, focusing on their new theory of moderate reasons-responsiveness as a model of "moral sanity." This theory, presented in *Responsibility and Control*, solves many of the problems with Fischer's earlier weak reasons-responsiveness model, such as its unwanted implication that agents who are only erratically responsive to bizarre reasons can be responsible for their acts. But I argue that the new model still faces several problems. It does not allow sufficiently for non-psychotic agents (who are largely reasons-responsive) with localized beliefs and desires incompatible with full responsibility. Nor does it take into account that practical "fragmentation of the self" over time may also reduce competence, since moral sanity requires some minimum level of narrative unity in our plans and projects. Finally, I argue that actual-sequence accounts cannot adequately explain sane but weak-willed agency. This is because without libertarian freedom, such accounts have no way to model the perverse agent's determination to be irrational or weak.

KEY WORDS: *akrasia*, compatibilism, control, free will, Harry Frankfurt, insanity, John Fischer, libertarianism, Mark Ravizza, Michael Bratman, moral responsibility, narrative unity, Peter van Inwagen, reasons-responsiveness, sanity, self, weakness of will

This paper is about the conditions of what I call "moral sanity," meaning the minimum capability to respond to practical reason that an agent must satisfy to be held morally blameworthy or praiseworthy for some element of their agency (e.g., an action, omission, or decision). Just as a legally insane agent cannot be held legally responsible, a morally insane agent cannot be held morally responsible. I argue that the most thoroughly worked-out recent attempt to explain moral sanity without libertarian freedom has many problems, the most serious of which is its difficulty in accommodating our normal experiences of "weakness of will."

In *Responsibility and Control*,[1] John Martin Fischer and Mark Ravizza make great strides in developing the semi-compatibilist account of moral responsibility first introduced in Fischer's article on "Responsibility and

[1] John M. Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998).

The Journal of Ethics **6:** 235–259, 2002.

Control,"[2] and further sketched out in Fischer's book, *The Metaphysics of Free Will*.[3] A good part of *Responsibility and Control* is devoted to arguing that simultaneous and pre-emptive over-determination cases (first introduced by Harry Frankfurt) show that moral responsibility for actions, intentions, omissions, and their consequences does not require libertarian freedom or "regulative control," as the authors call it (i.e., the freedom, in some circumstance *C*, to bring about a different intention, action, omission, or consequence than we actually do in *C*). In this article, I will focus instead on the authors' positive account of moral responsibility in terms of what they call "guidance control," which depends solely on features of the actual sequence of events and intentional states (including their dispositional properties) that causally explain the intention, action, omission, or its consequences.

1. Introduction: The Problems

In Chapters 7 and 8 of *The Metaphysics of Free Will*, Fischer suggested that we have guidance control over action *A* if and only if *A* was actually caused by a "weakly reasons-responsive mechanism," meaning one that leads us to do otherwise than *A* in some other possible world in which a mechanism of the same kind operates, and there is sufficient reason for us to do otherwise than *A*.[4] In making sense of this, it helps to imagine the "mechanism" as the process of intentional states – desiderative, emotional, evaluative, and so on – that the agent goes through in considering the circumstances and forming an intention to act. In telling this intentional story to make sense of an action, we might also refer to dispositions and habits of character, as well as longer-term intentions and commitments of the agent. All these could constitute mechanisms, or parts of a larger mechanism, in Fischer and Ravizza's sense. When an action *A* derives from a weakly reasons-responsive psychological mechanism, *A* is rationally guided in at least the

[2] John M. Fischer, "Responsibility and Control," *The Journal of Philosophy* 89 (1982), pp. 24–40, reprinted in John M. Fischer (ed.), *Moral Responsibility* (Ithaca: Cornell University Press, 1986), pp. 174–190. Also see John M. Fischer, "Responsiveness and Moral Responsibility," Ferdinand Schoeman (ed.), *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987), pp. 81–106.

[3] John M. Fischer, *The Metaphysics of Free Will* (London: Blackwell, 1994), pp. 160–189.

[4] Fischer, *The Metaphysics of Free Will*, p. 166. Also see Fischer and Ravizza, "Introduction" to John M. Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press, 1993), pp. 29–32.

minimum sense necessary for it to be imputable to the agent, or for it to be an appropriate target of reactive attitudes – so Fischer proposed in 1994.[5]

Several sorts of objections were raised against this preliminary analysis. (1) First and foremost, the condition of "weak reasons-responsiveness" is *too* weak: it allows intuitively non-accountable agents who are responsive to reasons in bizarre or haphazard ways to count as responsible.[6] (2) Second, on responsibility for consequences, Peter van Inwagen defended his argument that in cases where preemptive or simultaneous conditions make a consequence of an action or omission inevitable, the agent is responsible for the concrete event-particular, which she can avoid, while she is not responsible for the consequence-universal that she cannot avoid.[7] (3) Third, despite Fischer's past critiques and Ravizza's counterexample to van Inwagen's "direct" argument for the incompatibility of determinism and moral responsibility,[8] some philosophers still pursued this defense of incompatiblism.[9] (4) Fourth, several writers objected (as Fischer anticipated) that a reasons-responsive mechanism model of responsibility fails

[5] Note that in *The Metaphysics of Free Will*, p. 206, Fischer describes his sketch as "a first approximation to an account of moral responsibility for actions." And although his account there included new thoughts about how to specify the relevant reasons-responsive mechanism, it otherwise did not differ much from the model sketched in his earlier paper on "Responsibility and Control," or his Introduction (with Ravizza) to *Perspectives on Moral Responsibility*, pp. 31–32, in which (as in *The Metaphysics of Free Will*) most of the problems listed below were already acknowledged as requiring the further answers now presented in *Responsibility and Control*.

[6] See the following: Peter van Inwagen's example of a madman who is weakly responsive to reasons, in Peter van Inwagen, "Fischer on Moral Responsibility," *The Philosophical Quarterly* 47 (1997), p. 380; R. J. Wallace's objection that "What matters is not the ability merely to respond grasp and respond to (some) practical reasons, but the ability to grasp and respond to specifically *moral* reasons," in R. J. Wallace, *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press, 1996), p. 189; Ferdinand Schoeman's Saber-slayer example (mentioned in Fischer, *The Metaphysics of Free Will*, p. 243, note 8 and Fischer and Ravizza, *Responsibility and Control*, p. 65); and **Mark Ravizza's doctoral thesis, *Moral Responsibility and Control: An Actual Sequence Approach*, Ph.D. dissertation, Yale, 1994.**

[7] Peter van Inwagen first presented this argument in Peter van Inwagen, "Ability and Responsibility," *The Philosophical Review* 87 (1978), pp. 201–224, reprinted in Fischer (ed.), *Moral Responsibility*, pp. 153–173; he developed the argument further in Peter van Inwagen, *An Essay on Free Will* (Oxford: Oxford University Press, 1983); and he defended it again in van Inwagen, "Fischer on Moral Responsibility."

[8] See Peter van Inwagen, "The Incompatibility of Responsibility and Determinism," in M. Bradie and M. Brand (eds.), *Bowling Green Studies in Applied Philosophy*, Vol. 2 (Bowling Green: Bowling Green State University Press, 1980), pp. 30–37, reprinted in Fischer (ed.), *Moral Responsibility*, pp. 241–249.

[9] For example, see Ted A. [Fritz] Warfield, "Determinism and Moral Responsibility are Incompatible," *Philosophical Topics* 24 (1996), pp. 215–226.

because such a mechanism could have responsibility-undermining *sources*, e.g., direct manipulation of the brain, or being induced by hypnosis, programming, or conditioning.[10] Fischer and Ravizza try to resolve all these problems in *Responsibility and Control*, with many interesting results.[11]

These four issues concern the merits (i.e., conceptual cogency and phenomenological adequacy) of the actual-sequence model as a positive account of responsibility in its own right, quite aside from the further question of whether Frankfurt-inspired over-determination cases provide good grounds for *rejecting* "regulative control" or libertarian formulations of the freedom required for moral responsibility. Fischer and Ravizza reaffirm that libertarian freedom is incompatible with causal determinism,[12] but they deny that this freedom is required for moral responsibility. I have

[10] See Fischer, *The Metaphysics of Free Will*, p. 209; Eleonore Stump, "Intellect, Will, and the Principle of Alternate Possibilities," in Michael Beaty (ed.), *Christian Theism and the Problems of Philosophy* (Notre Dame: University of Notre Dame Press, 1990), reprinted in *Perspectives on Moral Responsibility*, pp. 237–262, pp. 258–261; Eleonore Stump, "Persons: Identification and Freedom," *Philosophical Topics* 24 (1996), pp. 189–191; and David Zimmerman, "Acts, Omissions, and 'Semi-Compatibilism,'" *Philosophical Studies* 73 (1994), pp. 209–223.

[11] There has also been an important debate, which is *not* rejoined in *Responsibility and Control*, over whether even from an incompatibilist viewpoint, such as an agent-causal view accepting the incompatibility of physical determinism and free will, Frankfurt-type cases show that an agent can be responsible for decisions she could not avoid. Fischer urged in "Responsibility and Control" that counterfactual intervener examples do not require that the actual sequence leading to the agent's action be a deterministic sequence: this is important, because otherwise such examples would seem to *presuppose* the truth of determinism. Others have argued that in cases of preemptive overdetermination of an action *A* or decision *D*, in the actual sequence there must be some event *E* in the agent that entails *A* or *D*, whose absence in the counterfactual sequence "triggers" the counterfactual intervener, and without this intervention, the absence of *E* entails that the agent does not do *A*, or does not decide *D*. Otherwise, the agent could do *A* or decide *D* without the intervener having assurance of it, and could avoid *A* or *D* without giving the signal upon which the intervener compels him to do *A* or *D*. See David Widerker, "Libertarian Freedom and the Avoidability of Decisions," *Faith and Philosophy* 12 (1995), pp. 113–118; John M. Fischer, "Libertarianism and Avoidability: A Reply to Widerker," *Faith and Philosophy* 12 (1995), pp. 119–125; David Widerker and Charlotte Katzoff, "Avoidability and Libertarianism: A Response to Fischer," *Faith and Philosophy* 13 (1996), pp. 415–421; David P. Hunt, "Frankfurt Counterexamples: Some Comments on the Fischer-Widerker Debate," *Faith and Philosophy* 13 (1996), pp. 395–401; Eleonore Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," in Jeff Jordan and Daniel Howard-Snyder (eds.), *Faith, Freedom, and Rationality* (Lanham: Rowman and Littlefield, 1996), pp. 73–88.

[12] Fischer and Ravizza, *Responsibility and Control*, pp. 17–25.

addressed some of these arguments in a separate review of their book,[13] but I leave them aside here, since even if semi-compatibilist arguments convince many philosophers that libertarian freedom is not essential to moral responsibility, it remains to be seen how well the semi-compatibilists can do in providing their own alternative account of its conditions. Moreover, even libertarians who reject semi-compatibilism may find many features of Fischer and Ravizza's account helpful for understanding moral sanity.

In addition, I will also leave aside issues (2), (3), and (4) for separate treatment in other articles. This means bracketing Fischer and Ravizza's innovative treatment of responsibility for omissions and consequences, their refutation of the direct argument for the incompatibility of responsibility and determinism, and their new Strawsonian account of "owning" or taking responsibility for the psychological mechanisms through which we control our actions and their consequences. Instead, I will focus on their account of reasons-responsiveness in the psychological mechanisms on which we act, which is the core of their theory.

2. Moral Sanity as Moderate Reasons-Responsiveness

The idea that responsibility requires that one's actions issue from reasons-responsive mechanisms is best understood, I think, as an interpretation of the twin cognitive and motivational aspects of *moral sanity*:[14] (a) that agents can *recognize* appropriate considerations as practical reasons for

[13] John J. Davenport, "Review of *Responsibility and Control*," *Faith and Philosophy* 17 (2000), pp. 384–395.

[14] Compare Susan Wolf's analysis in Susan Wolf, "Sanity and the Metaphysics of Responsibility," in Ferdinand Schoeman (ed.), *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987), pp. 46–62. In this essay, Wolf develops the notion of a sane "deep self," which is able to correct itself in a reasons-responsive fashion (p. 58). Note that sanity as a condition in which one's beliefs and values can be controlled by "perceptions and sound reasoning that produce an accurate conception of the world" (p. 55) is clearly an *epistemic* condition for responsibility (though it is surely not the only epistemic condition for responsibility for particular actions). Although guidance control involves more than this sane epistemic relation between objective reasons there are in the world and subjective reasons the agent can recognize, this still shows that guidance control is also a *partially* epistemic concept. As the analysis of its reasons-responsiveness component progresses, it becomes harder to say that "guidance control" is only an explication of the "volitional" side of imputable action, or of what it means to act voluntarily or "freely" in the sense required for responsibility. For guidance control clearly includes robust epistemic conditions (as becomes even more apparent when the authors treat the "taking responsibility" component of guidance control). Nevertheless, some other epistemic conditions must remain outside the definition of sane action, or sane agents could

them to act, and (b) that agents can *be motivated* by such practical reasons to act as they indicate. Failure to meet these conditions explains why agents are not responsible in the sort of cases that posed problems for earlier unsophisticated compatibilist models: e.g., persons whose choices follow their beliefs, but whose beliefs and values are so systematically distorted as to have little relation to reality, and persons who do what they want, but whose desires result from coercion, compulsive disorders, psychotic fears, brainwashing, or other forms of responsibility-undermining obsession, manipulation, or conditioning. The mechanisms on which such agents act in the actual sequence will usually fail to have the kind of reasons-responsiveness we think all sane agents must exhibit if they are to be held accountable, even when they knowingly do wrong.

Fischer and Ravizza are clear that "sanity" in the sense of appropriate reasons-responsiveness is not *all* that is required for moral responsibility. Moral sanity is only a necessary rather than a sufficient condition for moral responsibility, since an agent will not be responsible for an act within her guidance control if she fails to meet other epistemic conditions (e.g., appropriate knowledge of the circumstances of action). Moreover, even the act's being within *her* control involves more than its being morally sane, since a sane sequence of intentional states in the agent's mind could still be the result of more subtle forms of manipulation or intervention that would *alienate* the agent from them. For her to be responsible for them, the sane intentional mechanisms on which she acts must also be attributable to her, or "autonomous" (in one sense of this concept).[15] Fischer and Ravizza describe this condition in passive terms as

never be excused for reasons of nonculpable ignorance. I am indebted to an anonymous referee of *The Journal of Ethics* for this point.

[15] I say "in one sense" because I think there are phenomenologically distinct levels of "ownership of" or "identification with" an action, or with the intention so to act, or with the motives behind the decision to form the intention. These different levels create quite *distinct* senses in which an act, intention, or motive can be said to be "self-determined" or "autonomous" for the agent. In the most minimal sense (zeroth-level ownership), to say of any psychic mechanism or state S that it *belongs* to agent A is just to say that S is a functioning aspect of A's consciousness, A's sequence of intentional states, or A's psychē as a whole. But even a compulsive desire implanted against our will is "ours" in this minimal sense. In the next stronger sense (first-level ownership), a psychic state is "ours" or "autonomous" or "internal to us" if it has the authority of our will in it – or it bears our agent-imprimatur – even if the authority is not unequivocal, or our identification with the psychic state is not unqualified. This is the level Fischer and Ravizza try to explain in their account of taking responsibility: the agent recognizes the action and its motives as expressing her will. There may be even stronger senses of autonomy (as second-level ownership) involving wholehearted or unqualifiable commitment to the psychic state or the acts that issue from it. This is one way of understanding what Harry Frankfurt has called "decisive identification."

the mechanism being the "agent's own"[16] or in active terms as her having "taken responsibility" for it.[17] But they postpone an account of this aspect of "guidance control" until the last chapter, since explaining the kind of reasons-responsiveness that moral sanity requires is an important problem in its own right. Moral sanity is thus one of two main components of "guidance control" over our actions.

Four conditions for moral sanity. To understand the authors' account of moral sanity, we must note how the notion of reasons-responsiveness works in tandem with three other points. While (i) the agent must act on a suitably reasons-responsive mechanism, (ii) it must also be the *relevant* mechanism underlying the action that is sufficiently reasons-responsive. To pick out the relevant mechanism, we have to ignore ones that are "temporally extrinsic," or *defined* by relation to a specific action-outcome (analogously to "soft facts"). The relevant mechanism must be temporally intrinsic: it must not be specified so as to *entail* the action it explains.[18] In addition, the relevant mechanism must be the one in terms of which we would intuitively explain the action. Fischer and Ravizza assume that "for each act, there is an intuitively natural mechanism that is appropriately selected as the mechanism that issues in an action, for the purposes of assessing guidance control and moral responsibility."[19] The idea seems to be that a "mechanism" is like an intentional explanation: it must be possible to give some coherent narrative in terms of reasons, motives, considerations, etc., that make intelligible why A did X. This could be fleshed out in terms of what the agent's own honest and reflective account of her act would be, or (to deal with unconscious factors) in terms of what an ideally informed observer or omniscient judge of minds would say about the action. Nevertheless, specifying the relevant psychological mechanism remains problematic in much the same way as specifying the relevant maxim behind an action in Kantian ethics.

I note in passing that it seems unnecessary to make the strong assumption that in every case, there can only be *one* actual-sequence mechanism that explains a given action (or omission). Sometimes it is natural to point to several intuitively "different" mechanisms at work simultaneously in the same agent, which are all relevant for the particular action in question. The authors may fear that this would open their theory to actual-sequence over-determination objections. But such cases are the norm, rather than

[16] Fischer and Ravizza, *Responsibility and Control*, pp. 38–39.

[17] Fischer and Ravizza, *Responsibility and Control*, pp. 207–239.

[18] Fischer and Ravizza, *Responsibility and Control*, p. 46.

[19] Fischer and Ravizza, *Responsibility and Control*, p. 47.

being unusual. For example, it would not be at all surprising if multiple reasons operated in a given agent to cause him to read a certain novel: he heard it was entertaining; he wants to improve his reading; he read a review which implied that the book can educate him about important historical topics; and (by chance) he is irrationally attracted to the picture on the book's cover. Surely he is still responsible for reading the book (perhaps he deserves mild praise for it), although at least one non-reasons-responsive factor contributed to his reading it. The only plausible solution may be to say that whether a person exercises the guidance control required for moral responsibility can be ascertained only by a prudential judgment about the circumstances behind the action: it involves an interpretation concerning what mechanism(s) were *most important* in the causal or intentional account of an agent's action. In other words, a kind of psychological hermeneutics is unavoidable in specifying the intentional mechanism(s) on which the agent acted, and interpreting which ones were most relevant to the action.

(iii) Third, in judging whether the mechanism(s) on which an agent acts have the "dispositional property" of reasons-responsiveness,[20] we must consider what happens in alternative scenarios where the *same kind* of mechanism operates, but there are sufficient reasons to do otherwise than the agent did in the actual sequence. But "sameness of kind" here does not require that the mechanisms in the actual and counterfactual sequences are identical in every detail, down to the micro-level. Thus the relevant mechanism(s) need not be defined in terms of a particular causal process in the brain that would nomologically necessitate the action, if causal determinism holds. There only need to be scenarios in which the same *sort* of mechanism – not the same mechanism-particular – operates and the agent follows different reasons and does otherwise.

(iv) Fourth, the authors add a *tracing condition* to their account: a person can be responsible for acting on a non-reasons-responsive mechanism if and only if that mechanism itself issued from a suitably reasons-responsive process earlier in the actual sequence. For example, the mechanism on which one acts when drunk may not be suitably reasons-responsive by itself, but if one had guidance control in getting drunk (or further back, in acquiring the bad habit), then one is responsible for actions like drunk driving and their consequences.[21] The authors admit that this is only a sketch of the "tracing" principle needed to deal with such cases, but they do not consider more complex examples. It would be interesting to see this worked out more fully, because as I have argued

[20] Fischer and Ravizza, *Responsibility and Control*, p. 53.

[21] Fischer and Ravizza, *Responsibility and Control*, pp. 49–50.

elsewhere, libertarians can make use of very similar tracing conditions to blunt the intuitive force that Frankfurt-style overdetermination cases have against simple (or untracing) libertarian conditions on moral responsibility.

Strong and weak reasons-responsiveness. In Chapter two, Fischer and Ravizza reconsider the two kinds of reasons-responsiveness distinguished in Fischer's *The Metaphysics of Free Will* and earlier works: namely, "weak reasons-responsiveness" (WRR) and "strong reasons-responsiveness" (SRR). The latter requires that *if* a mechanism of the same kind on which the agent acts in the actual sequence were to operate in an alternative scenario, "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."[22] SRR rules out any failure to recognize relevant reasons for actions, a failure "typically associated with delusional psychosis."[23] It also rules out failure to react properly to recognized reasons, a problem that "afflicts certain compulsive or phobic neurotics."[24] It also rules out agents who fail to translate their choices into actions, e.g., because of physical incapacities.[25]

But SRR is too strong a requirement for responsibility: it requires too tight a fit "between the reasons there are, and the reasons the agent has, the agent's reasons and his choice, and his choice and action."[26] For SRR also rules out cases of weakness of will, where the agent recognizes a sufficient reason not to do something, and does it anyway, or cases where the agent chooses (or forms an intention) to do what she recognizes she has sufficient reason to do, but then does not do it when the time comes.

The authors instead propose that responsibility only requires something like the looser fit between reasons, choices, and actions defined by acting on a weakly reasons-responsive mechanism, which only requires that "there exist *some* possible scenario (or possible world) in which there is a sufficient reason to do otherwise, the agent recognizes this reason, and the agent does otherwise."[27] For weak-willed persons can satisfy WRR: "Even

[22] Fischer and Ravizza, *Responsibility and Control*, p. 41.

[23] Fischer and Ravizza, *Responsibility and Control*, p. 41.

[24] Fischer and Ravizza, *Responsibility and Control*, p. 42.

[25] Fischer and Ravizza, *Responsibility and Control*, p. 42.

[26] Fischer and Ravizza, *Responsibility and Control*, p. 42.

[27] Fischer and Ravizza, *Responsibility and Control*, p. 44. In WRR, it is the fact that there is *some* such world (sharing the same laws), rather than its similarity to the actual world, that counts. The opposite is true for SRR, which is satisfied only if in *any* world – including those closest to the actual world – in which the mechanism operates, if a similar choice-circumstance occurs, and yet the reasons for action dictate a different action, the agent acts as reason dictates and does otherwise than in the actual sequence.

an agent who acts against good [or sufficient] reasons can be responsive to *some* reasons."[28] In other words, agents can be morally sane without being strongly reasons-responsive. By contrast, if an agent fails to be even weakly reasons-responsive, this is a sign of a responsibility-undermining compulsion or some similar disorder: e.g., if someone would persist in stealing a book despite the anticipation of *any* consequence no matter how horrible (even the death of his whole family), then "the actual mechanism would be inconsistent with holding him morally responsible for his action."[29] Similarly, an agent whose action is determined by physical processes caused by a physiological addiction to some drug acts on a literally irresistible urge, and this physical process is certainly not weakly reasons-responsive.[30] The agent may know better, but she *cannot* conform her actions to her reasons: they are ruled by a non-responsive mechanism.

Moderate reasons-responsiveness. Fischer and Ravizza then introduce the notion of "moderate reasons-responsiveness" (MRR), which includes several modifications designed to avoid problems with the weak reasons-responsiveness model. First, when an agent acts on a moderately reasons-responsive mechanism, the agent would do otherwise in at least some of the possible scenarios in which there is sufficient reason to do otherwise, and would "do otherwise *for that reason*."[31] This caveat is meant to require an appropriate relationship between the action and the cognitive grasp of reasons that rules out deviant causal chains (but the authors do not try to specify that relationship further).

Second, in a moderately reasons-responsive mechanism, the "fit between reasons and actions" is tighter than in WRR, though still looser than in SRR.[32] Building on Bernard Gert and Timothy Duggan's idea that a responsible agent must be disposed to "an appropriate pattern of responses" to different incentives or reasons to act,[33] Fischer and Ravizza propose four conditions for moderate reasons-responsiveness: (1) The psychological mechanism on which the agent acts must be *regularly receptive* to reasons to act otherwise, which the authors interpret as meaning that "it involves an *understandable pattern* of (actual and hypothetical) reasons-receptivity,"[34] and (2) the agent's subjective reasons are

28 Fischer and Ravizza, *Responsibility and Control*, p. 45.
29 Fischer and Ravizza, *Responsibility and Control*, p. 45.
30 Fischer and Ravizza, *Responsibility and Control*, p. 48.
31 Fischer and Ravizza, *Responsibility and Control*, p. 64.
32 Fischer and Ravizza, *Responsibility and Control*, p. 68.
33 Fischer and Ravizza, *Responsibility and Control*, pp. 66–67.
34 Fischer and Ravizza, *Responsibility and Control*, p. 71.

"at least minimally 'grounded in reality.' "[35] Someone who, for example, would recognize a bribe of $1000 as a reason not to try an appealing drug, but would not be stopped by a price of $2000 or more,[36] seems to be judging in the sort of confused, erratic, or idiosyncratic ways that characterize various sorts of mental illness or loss of cognitive function. Likewise, someone who recognizes a coherent pattern of reasons, but derives them from delusions and hallucinations is not morally sane, and thus lacks guidance control.

(3) The psychological mechanism on which the agent acts must be at least *weakly reactive* to reasons subjectively recognized.[37] The authors think a stronger requirement for reactivity would tend to rule out "weakness of will" again, and is not necessary because "reactivity is all of a piece in the sense that the mechanism can react to all incentives [to do otherwise] if it can react to one."[38] Hence an agent need only demonstrate "*some* reactivity, in order to render it plausible that his mechanism has the 'executive power' to react to the actual incentive to do otherwise" in the case for which we are holding him responsible[39] (I critique this idea in the last section of the paper).

(4) Finally, the authors modify their earlier accounts so that an appropriate subject of moral judgments and reactive attitudes must acts on mechanisms that are moderately receptive to *moral* reasons, as well as other sorts of reasons. Some smart animals, young children, and psychopaths may regularly recognize and react to *some sorts* of relevant reasons to act otherwise, but still fail to be moral agents, because they do not recognize or react to moral reasons, as distinct from merely pragmatic or instrumental reasons.[40] Yet, while the authors require morally responsible agents to be receptive to at least some "understandable pattern" of moral reasons (as understood in their community),[41] they did not explicitly require that such agents be even weakly *reactive* to moral reasons.[42]

For convenience, the main features of Fischer and Ravizza's semi-compatibilist theory of the conditions for responsible agency can be summarized in the following compressed form:

35 Fischer and Ravizza, *Responsibility and Control*, p. 73.
36 Fischer and Ravizza, *Responsibility and Control*, p. 70.
37 Fischer and Ravizza, *Responsibility and Control*, p. 73.
38 Fischer and Ravizza, *Responsibility and Control*, p. 74.
39 Fischer and Ravizza, *Responsibility and Control*, p. 75.
40 Fischer and Ravizza, *Responsibility and Control*, pp. 76–77.
41 Fischer and Ravizza, *Responsibility and Control*, p. 77.
42 Fischer and Ravizza, *Responsibility and Control*, p. 79.

Semi-compatibilist conditions for moral responsibility

I Epistemic conditions: the agent cannot be (nonculpably) ignorant of the circumstances of his action.

II Freedom/Control conditions: the agent cannot have been (non-culpably) forced to do what he did. His action must count as originating from him, or be under his control or *self-determined.*

A. *Moral sanity*: the agent must be appropriately reasons-responsive in deciding how to act:

1. The agent acts on (at least one) moderately reasons-responsive psychological mechanism.

   (i) (cognitive) The mechanism must be regularly *receptive* to reality-tracking reasons to act otherwise, including moral considerations: these reasons must form at least a somewhat coherent pattern (understandable by an external observer) and be somewhat grounded in reality. In other words, the agent must be capable of recognizing a regular pattern of considerations as reasons for him to act one way rather than another.

   (ii) (motivational) The mechanism must be at least weakly *reactive* to practical reasons it recognizes: this means that in at least one of the possible worlds where the agent recognizes sufficient reason(s) to do otherwise than in the actual sequence, he would do otherwise as the reason(s) prescribe, choosing his different act for those reason(s).

2. The relevant MRR-mechanism must be temporally intrinsic (does not entail the act), and must be the one that an informed interpretation would specify as most important in an intentional explanation of the action.
3. Reasons-responsiveness as a dispositional property of a mechanism *M* is measured by considering what happens in possible worlds where the *same kind* of mechanism operates in the agent, but not necessarily the very same mechanism-particular.
4. Or if (1) is not satisfied, then the agent's act *X* satisfies a tracing principle: the mechanism *M* behind *X* was itself developed or predictably caused by prior actions satisfying (1).

B. *Autonomy*: The agent takes responsibility for the psychic processes or intentional mechanisms behind her action, or recognizes them as "her own" or as self-determined in a suitable sense.

This summary helps the reader see clearly how the conditions of moral sanity fit into the entire account of moral responsibility.

3. Five Problems with the Formulation of MRR in *Responsibility and Control*

In my judgment, these amendments do much to alleviate the recognized problems with the earlier model, but lingering difficulties remain with the new model of moderate reasons-responsiveness. First, Fischer and Ravizza need to modify their conditions for MRR to require moderate reactivity to recognized moral considerations.[43] Otherwise, they will have to count as responsible an agent like Susan, who can recognize that objective moral considerations give rise to subjective reasons for her to act, but whose actions cannot be controlled by such recognition. Susan cannot be called morally weak-willed, since she cannot conform her actions to her moral sense at all. This may not be because of any compulsion: she may act on psychological mechanisms that are responsive to other kinds of nonmoral reasons. But there seems to be something wrong with treating her as a responsible agent, since she lacks the capacity for *morally guided* control altogether. She may be receptive, but cannot be said to be "responsive," to "a range of reasons that include *moral* reasons."[44] But a morally sane agent must be responsive to such reasons as well. Here a comparison to legal sanity is instructive: a legally sane agent must not only be able to tell the difference between right and wrong at the time of acting (which is all that United States law requires); she must also be able to conform her intentions to this knowledge.[45]

The second problem can also probably be addressed with a simple modification or clarification. As I read it, MRR is still compatible with some consistent but highly abnormal patterns of subjective reasons or incentives for action, when these are a localized subset of a generally reality-guided doxastic framework. Suppose that Malcolm is a psychologically normal forester, with no prior religious beliefs but no strong atheistic convictions either. One night, Malcolm has a dream of unusual strength and clarity, in which a nature spirit appears to him saying that when a full moon falls on Friday the 13th, the gods require a human sacrifice. Normally, his own mechanisms of rational doubt concerning dreams would lead him to reject the notion that a dream could be a religious revelation. But he has the dream several more times, and slowly begins to believe it. Maybe it is because of his brief stint with an animist cult ten years ago; maybe it is from watching too many bad horror movies, but Malcolm

[43] In his "Reply to Critics" at the Pacific APA in Albuquerque, NM (April 2000), Professor Fischer suggested since reactivity is "of a piece," an MRR mechanism will be reactive to all the reasons it recognizes, including moral ones. Note that this seems to revise footnote 23 in Fischer and Ravizza, *Responsibility and Control*, p. 79.

[44] Fischer and Ravizza, *Responsibility and Control*, p. 81.

[45] Thus the American Bar Association's model legal code calls for distinguishing the "cognitive" and "volitional" component of legal sanity. This distinction was made in *United States vs Freeman* (1969), but later overturned by an act of the United States Congress after a jury using this standard ruled John Hinkley insane during his attempted assassination of former U.S. President Ronald Reagan. This remains a deep problem in our working legal definitions of moral sanity.

is not sufficiently reflective to consider possible unconscious sources of his dream. So almost without realizing it, Malcolm starts to modify his doxastic framework to incorporate the deviant belief from his dream. This requires fewer changes than one might think: for instance, believing that dreams of a special kind can sometimes reveal genuine commandments not meant to be shared with others. But otherwise Malcolm's belief set remains average and as reality-guided as the next person's: it simply has one unusual streak of beliefs running through it. Importantly, his belief set remains responsive to moral reasons; he knows that killing innocent persons is (almost) always wrong. But he believes this prohibition is outweighed by his religious duty to perform the sacrifice on rare occasions.

Three years later a full moon Friday the 13th finally occurs, and Malcolm tries to act on his belief. The mechanism on which he acts in this circumstance may be receptive and reactive to a set of reasons consistent with his beliefs: e.g., if his victim convinced him that the moon was not quite full tonight, he would stop the sacrifice. Moreover, the beliefs on which he acts are a consistent part of a coherent doxastic framework that is largely reasons-responsive (albeit severely distorted on one topic). So by Fischer and Ravizza's criteria, Malcolm would seem to be fully responsible for the murder. But given the effect of his dreams, Malcolm seems substantially different from a normal agent who does something wrong for an idiosyncratic but still intelligible reason, when she ought to have known (or did know) better. The reasons to which Malcolm is responsive in this case are bizarre enough that they verge on being unintelligible to others, and as a result the jury (if convinced that Malcolm's testimony was sincere) might well wonder if we can reasonably think Malcolm should have known better. A reasonable jury might consider his sanity too diminished for him to be guilty of murder in the first degree, at least.[46] The point is that this kind of localized insanity can occur without the psychological mechanisms on which the agent acts being completely unresponsive to a consistent pattern of reasons, including moral reasons, and without these being largely ungrounded in reality.

Similar problems will arise with agents who have very unusual desires and emotions that create consistent patterns of reasons for them to act in highly bizarre ways, which nevertheless do not constitute compulsions or

[46] I do not, of course, want to suggest that recent U.S. verdicts in prominent cases with insanity defenses are any measure of what is reasonable. For example, the conviction of Andrea Yates – a chronically depressed and borderline mother who killed her children – strongly suggests that our public may have reached the point where vindictiveness and outrage blocks any serious consideration of the fact that the offender was suffering from major depressive episodes which robbed moral considerations of its normal motivational force at the time of the crime.

psychoses (some cases of fetishism and paraphilia may qualify here). Likewise, even if she is responsive to some abstract moral reasons, we usually suspect a responsibility-reducing psychological problem in an agent who is completely uninterested in engaging others, wants no involvement in the practices of her community, or who even withdraws completely from all social interaction. Thus Theodore Kaczynski, the "Unabomber," may in some fashion have "known" that it is wrong to kill innocents, and his extreme anti-technological philosophy may still have been responsive to a range of moral reasons (even if it led him to misjudge who was innocent and guilty), but his pattern of preferences and desires is clearly suggestive of some personality disorder that probably diminishes responsibility. MRR as currently formulated does not provide sufficiently for such cases of moral or legal insanity without psychosis. Enjoying guidance control may require more *objective rationality* in personal preferences and desire than MRR suggests. Since this is also a problem with our current law,[47] moral philosophy needs to give critical direction on this point.

Third, Fischer and Ravizza's definition of the regular reasons-receptivity of a psychological mechanism in terms of its intelligibility to concerned *third parties* is also problematic. For outside observers can be deceived, and fail to see the inner logic or regularity in someone's subjective interpretation of reasons for action. If we say instead that a reasons-receptive mechanism would be intelligible to an *ideally informed observer*, we still run into the problem that we sometimes have reasons to act that are entirely inscrutable, that make sense to us as an inner prompting, but that could not even *in principle* be articulated in terms accessible to outside third parties. Sometimes reasons have a kind of essential self-reference, or relation to the whole of an individual's sense of his/her identity and place in the world, that makes them necessarily opaque to any other subjective perspective.[48] These can still be *practical reasons* functioning as part of sane action, even if their significance is felt only in an inchoate sense of what matters, or what questions are salient and what

[47] The still-used but grossly flawed and outdated McNaughten standard (8 English Rep. 722 1844; NB spelling varies) concerns only receptivity to moral reasons, but not reactivity to them. It is even further from recognizing insanity in cases where systematically distorted or disturbed doxastic and emotional frameworks prevent the proper functioning of the agent's ordinary receptivity and reactivity to moral reasons in other contexts.

[48] David Wiggins recognizes a similar point in his discussion of Peter Winch and Alasdair MacIntyre on situations where moral judgment depends on maximally specific circumstances, including an uniterably concrete agent point of view: see David Wiggins, "Truth, and Truth as Predicated of Moral Judgments," in David Wiggins (ed.), *Needs, Values, and Truths*, 3rd ed. (Oxford: Oxford University Press, 1998), pp. 139–184 and 169–171.

kinds of inferences we might draw from salient considerations.[49] Reasons that operate at this level, as part of our whole practical frame of reference, may not be linguistically expressible in terms accessible to third parties. Yet someone acting on a mechanism guided by such reasons could still be morally responsible; indeed he might be acting on the deepest sort of practical reason possible.[50]

This point connects with R. Jay Wallace's concern that Fischer and Ravizza's focus on the modal properties of mechanisms "brings an objectifying, third-personal vocabulary to bear on phenomena that have their natural place within the deliberative perspective of practical reason . . ."[51] Wallace's point follows Kant's view that the notion of moral responsibility is intelligible only from *within* the practical standpoint of agency. My suggestion is that this standpoint of agency is not even always deliberative in the sense of involving an articulable thought-process.[52] This does not mean that *any* prompting or intuition whatsoever can serve as a basis for sane action, though. To explain the competencies of morally responsible agents, we may have to acknowledge substantive limits on the *content* of their doxastic and motivational sets, and on how the considerations on which they act fit in as one part of the *narrative whole* of their practical orientation in life. Without trying to spell this out here, some minimally coherent conception of "the good" or "the meaningful," along with some level of motivation guided by it, might be necessary for moral sanity. There are two sub-questions here. The first concerns whether Fischer and

[49] Ronald de Sousa has suggested that emotions function in this way as patterns of salience that help us avoid the philosopher's framing problem for action. See Ronald de Sousa, *The Rationality of Emotions* (Cambridge: MIT Press, 1987), pp. 194–203. I think this problem is important because in some cases it bears on how we come to "care" about things in Frankfurt's sense, or how we determine "what to care about" in the process of building a meaningful life.

[50] This would seem to be Martin Heidegger's view in Martin Heidegger, *Being and Time*, trans. John Macquarrie and Edward Robinson (New York: Harper and Row, 1962), H52–H179, pp. 78–224, where he suggested that authenticity is measured by responsiveness to our whole "being-in-the-world," i.e., the totality of "involvements" or pratical significances, matterings, and salient values that make up the gestalt of the practical universe, or the personal world in which we act as agents.

[51] R. Jay Wallace, "Review of *The Metaphysics of Free Will*," *The Journal of Philosophy* 94 (1997), p. 159.

[52] For example, consider Henry Bugbee's claim that "It is of the essence of authentic commitment that it be grounded behind the intellectual eye and not merely in a demonstrable basis which we can get before us. The ultimate meaning of service lies just here: We cannot gain command of what grounds our actions"; see Bugbee, *The Inward Morning*, reprinted with a new introduction by Edward Mooney (Athens: University of Georgia Press, 1999), p. 69. If Bugbee's appeal to a radical kind of "heteronomy" sounds like Emmanuel Levinas here, it is probably because they owe a common debt to Gabriel Marcel.

Ravizza's criteria for moral sanity are too formal. The second concerns what Fischer and Ravizza call the "externalism" of their account, or their focus on the history of the agent's internal set of motives and its relation to the world.[53] The viability and coherence of an individual's internal "mental economy" *as a whole* may matter just as much for sanity, and this coherence can be lacking even if, *taken separately*, most parts of that mental economy have respectable causal pedigrees in moral society-based and world-guided belief- and motive-forming processes.

This problem with externalism links directly with my fourth criticism: MRR seems to take too local an approach to sanity. It suggests that an agent is sane – and hence acts voluntarily – in a particular instance if the mechanism on which she acted in that one instance was regularly receptive and at least weakly reactive to reasons (now including some moral ones) that do not entirely fail to "track reality." But what if, for the past five years, this person has only acted on compulsive psychological mechanisms that are not reasons-responsive, or on mechanisms that are only responsive to delusional reasons, and then suddenly she has a brief period of clarity – say, a one-day break in her paranoid schizophrenia, perhaps – in which she acts on a mechanism satisfying MMR, after which she unfortunately return to her delirium? Is one day of reasons-responsiveness enough to qualify her as responsible during that day?[54] Would such a violent change itself not be disorienting enough to produce all kinds of emotional tensions in the agent? Could she pick up where she left off, be motivated by past resolutions, or without difficulty start deliberating with an eye to the future?

Alternatively, consider a less extreme case. Suppose that the agent in question, Mitchell, is a disturbed fellow in his mid-twenties, who has no direction in life, and is desperately turning from one ideology or set of comforting answers to another. The psychological mechanisms guiding his decisions on important life-choices (on education and career, what to do with an inherited fortune, what relations to cultivate with significant others, etc.) alter radically in quick succession. Mitchell is enthusiastic about going to college, even consumed with planning his studies, but then nonchalantly drops out the second week, without giving it a second thought. After an intense romance of four weeks, he gets married, and then immediately divorces, but acts as if this is perfectly natural. After a week with the Moonies, he gives most of his fortune to this cult, but then passion-

[53] Fischer and Ravizza, *Responsibility and Control*, p. 252.

[54] If someone says "yes" to this question, I suspect this illustrates the cultural effect of our legal institutions and their theoretically inadequate conception of moral sanity, which encourages us to look at each act on a snapshot basis.

ately pursues a lawsuit to get it all back. He reads atheist literature with fervent intensity for a week, to the point of physical exhaustion, and then burns all the books, and goes off to join a monastery. After one month there he leaves to pursue a solitary life of mountain-climbing in the wilderness, but being bored, he quickly flies to New Orleans where he enjoys a week of completely uninhibited debauchery and drunkenness during Mardi Gras. And so on. At best, we would regard such an agent as highly neurotic; at worst, we would question his sanity, especially if (as we suspect) his violently shifting priorities result from a frequent and largely random set of *changes in the psychological mechanisms on which he acts*. Individually each of these mechanisms may be moderately reasons-responsive, but it seems to be a mistake to hold Mitchell responsible *piecemeal* for each of the associated acts by tracing it to its short-lived MRR mechanism. At least it seems evident to me that Mitchell's behavior is erratic enough to cast serious doubt on his *full* responsibility for his actions.[55]

This example suggests that full moral competency also seems to require at least some minimal degree of rational connectedness between one's action-controlling mechanisms over time, or some coherence in the pattern of change from one mechanism to another. This is because fully competent agents must have some capacity for commitment and staying-power. MRR as presently formulated does not capture this crucial feature of rational agency. Narrative accounts of sane agency, like the one offered by Alasdair MacIntyre, do better on this score.[56]

Here Michael Bratman's analysis of temporally extended commitment in terms of planning might offer one way to modify MRR to address this problem.[57] As Bratman says, "the plans of planning agents will normally have a certain stability, persist through time, and structure later

[55] One anonymous referee of *The Journal of Ethics* insisted that in this case, as in the others above, Fischer and Ravizza can simply say that the agent is responsible. My response is: any bullet can be bitten, but not without sacrificing phenomenological adequacy. I'm willing to hazard that most readers would share my doubts concerning Mitchell, Malcolm, and Susan, and at least want to ask them questions before assuming that they are fully responsible. Given my incomplete stories about them (and philosophical examples are necessarily always incomplete) the presumption seems to be *against* full responsibility in their cases. Kaczynski is real, not a philosophical fiction, but here one has to ask if the widespread belief that he was sane is really guided by the evidence, or if instead it derives from an inexcusable (although common) public desire for revenge.

[56] See Alasdair MacIntyre, *After Virtue: A Study in Moral Theory*, 2nd ed. (Notre Dame: University of Notre Dame Press, 1984), Chapter 15.

[57] See Michael Bratman, "Responsibility and Planning," *The Journal of Ethics* 1 (1997), pp. 27–43, reprinted in *The Faces of Intention* (New York: Cambridge University Press, 1999), pp. 165–184; and Michael Bratman, *Intentions, Plans, and Practical Reasons* (Cambridge: Harvard University Press, 1987).

conduct."[58] He argues, following Peter Strawson, that only planning agents will have the capacity for the sort of ordinary interpersonal relationships within which we regard others (and ourselves) as responsible agents.[59] Frankfurt's analysis of caring as a way in which the agent reflexively guides her own motivational states over time also seems relevant to this problem.[60] But since neither Bratman nor Frankfurt focus on minimum threshold conditions for responsible agency or "moral sanity," Fischer and Ravizza must decide what sort of capacity for thematic integration of purpose over time, or partial narrative unity, is essential for being an apt candidate for the reactive attitudes. They must also consider whether Bratman's, Frankfurt's, or some other model gives us an adequate understanding of the kind of sustained "will" or volitional commitment that is evidently lacking in Mitchell's erratic behavior.

Weakness of will. The fifth and final problem is the most serious in my judgment. I think "weakness of will" (at least in some of its varieties) poses *special* problems for actual-sequence models of moral responsibility, that is, problems *above and beyond* the difficulties admittedly posed by various kinds of *akrasia* for any theory of the freedom involved in moral responsibility. Suppose for the moment we follow Robert Dunn's suggestion that true weakness of will occurs when "an agent knowingly and intentionally act[s] against his full-fledged all-out summary better judgment, or judgment of what is right."[61] On libertarian accounts of responsibility that distinguish evaluative judgments, motivational attitudes, and intention-formation sufficiently to make weakness of will in Dunn's sense possible, the problem is usually to show that electing an option recognized to be inferior is not simply an arbitrary move, and to explain why the weak-willed act can still be regarded as (minimally) autonomous or self-determined. This is part of the larger libertarian problem of explaining how any election among options can have reasons (or some intentional story) behind it, without these just causally determining the selection among options.

As tough as these problems are for libertarians, I think the situation is worse for actual-sequence theories. In attempt to leave room for weakness of will in their model, Fischer and Ravizza argue that guidance

[58] Bratman, "Responsibility and Planning," p. 170.

[59] Bratman, "Responsibility and Planning," pp. 171–180.

[60] See Harry Frankfurt, "The Importance of What We Care about," in *The Importance of What We Care about*, p. 83; and "On Caring" in Harry Frankfurt, *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999), pp. 161–162.

[61] Robert Dunn, *The Possibility of Weakness of Will* (Indianapolis: Hackett Publishing Company, 1987), p. 1.

control requires only weak responsiveness of intentions and decisions to recognized reasons. Weakness of will then occurs when the mechanism *could but does not* respond to a recognized sufficient reason to do otherwise.[62] In defense of this analysis, Fischer and Ravizza argue that if the agent's psychological mechanism *M* can react to one recognized reason for acting otherwise (in an alternative scenario), then it can react to any such incentive or consideration recognized subjectively as a reason to do otherwise. Reactivity is in this sense "all of a piece."[63] This raises at least two problems.

(1) It is psychologically implausible to hold that just because a mechanism can react to one kind of recognized incentive for doing otherwise, it can react to *any* such incentive. Granted, it would be odd for my intentional process to be *responsive* to just one very narrowly specified kind of consideration, even though it is *receptive* to many others (this would be a kind of monomania). For instance, suppose the psychological mechanism on which I act in making some political decision is envy (in the Rawlsian sense of a desire to reduce merely relative differences in holdings, even at the price of leveling). Suppose we stipulate that this mechanism (*E*) would be responsive to new evidence that relative inequality in my community is not as great as I thought: on coming to believe this, I would adopt a different course of political action. Then most likely *E* would also be responsive to evidence that the rich hope to increase unjust advantages to their offspring by repealing all inheritance taxes. But this flexibility (say, to agitate for more or less radical redistributive policies) need not extend to all considerations that I recognize as relevant. This envy-mechanism could be stubbornly unreactive to my friend's cogent argument that envy is a morally suspect attitude (even though I am receptive to this argument, and perhaps somewhat ashamed of being envious). So why should we think that a practical mechanism that is flexible in one respect should be flexible in *all* the other ways allowed by considerations to which the agent is receptive?

It is not clear, then, that a psychological mechanism normally has the power to react to (or be guided by) *all* the reasons that it (or its agent) can recognize as reasons for the agent to do (or refrain from) something. In other words, it will not usually be true that for every reason *R* that mechanism *M* recognizes as a sufficient reason to act otherwise than one does in the actual sequence, there is some possible world in which *M* acts on *R*. Fischer and Ravizza suggest that if there is an *R* such that in no world wherein *M* recognizes *R* does it act on *R*, then *to act* on

[62] Fischer and Ravizza, *Responsibility and Control*, p. 42.

[63] Fischer and Ravizza, *Responsibility and Control*, p. 73.

R would be to act on "a different mechanism."[64] But this strong claim seems less convincing than the following weaker alternative: namely, that psychological mechanisms could be individuated by the set of subjectively recognized reasons or incentives to which they react across the range of possible worlds in which they operate. On this weaker thesis, a mechanism would be able to react to all the sorts of reasons or incentives that are included in its "essential set": to act on other reasons outside this set would necessarily be to act on a different mechanism. But this would not entail the stronger thesis that *M* cannot even *recognize* or be receptive to a reason outside its essential set, i.e., one to which it is incapable of reacting (or by which is cannot be guided). In my example, the process of intentional states constituting "envy" may well include cognitive recognition that there is something wrong with a desire to reduce relative differences at any cost, but without this recognition causing any related motivation not to gratify envy-desire if possible. Sometimes psychological mechanisms are motivationally disengaged from considerations whose cognitive force they nevertheless apprehend. Or if Fischer and Ravizza reject this, then at least they owe us further explanation about how evaluative judgments and motivational attitudes can be connected and disconnected within a single psychological mechanism, or as parts of a single discernible intentional process leading to action or inaction.

(2) Even if this thorny problem about the reactivity of mechanisms can be solved, Fischer and Ravizza face a further dilemma regarding the analysis of *akrasia* in terms of the dispositional properties of mechanisms. This dilemma arises because a phenomenologically adequate theory of responsibility for weak-willed decisions and actions must explain both the following conditions:

(1) *Freedom*: in what sense the better and the worse option(s) are available to the agent; and
(2) *Agent-weakness*: how the practical irrationality or perversity involved in taking the worse option is attributable to the agent (or in what sense she chooses this option *as worse*, or *qua inferior*).

Fischer/Ravizza-style actual-sequence accounts seems to explain the (a) condition in a way that rules out an adequate explanation of (b). They want to say, *loosely* speaking, that the agent is weak-willed because the weakly reasons-reactive mechanism *M* on which she acts in the actual sequence "could have responded" to a reason to do otherwise, but did not. But *strictly* speaking, their account only says that the better option is available in the sense that *M* has a certain dispositional property *D*: in

[64] Fischer and Ravizza, *Responsibility and Control*, p. 74.

some possible worlds, *M* is guided by reason *R*, which justifies forming the better intention. They add that since *R* is also apprehended in the actual sequence, the better option was available here too:

> . . . a mechanism's reacting differently to a sufficient reason to do otherwise in some other possible world shows that the same kind of mechanism can react differently to the *actual* reason to do otherwise. This general capacity of the agent's actual sequence mechanism – and *not* the agent's power to do otherwise – is what helps to ground moral responsibility [in cases of weakness of will].[65]

Yet how exactly do we understand this phrasing that *M* "can react differently" to reason *R*, which is a reason to do otherwise (and better) than it does in the actual sequence? The strictly dispositional account tells us to read this as follows: *M* has the modal property *D*, namely that it reacts to *R* in at least one *R*-world, i.e., worlds in which it recognizes *R* as a sufficient or overriding reason to act. This reading is of course compatible with complete psychophysical determinism: *M* can have property *D* even if it turns out that all *R**-worlds (i.e., those on which *M* acts on *R*) are inaccessible to the agent for other reasons (e.g., because of interveners, simultaneous over-determination, etc.).[66] But Fischer and Ravizza seem to want something more than this. In the passage just quoted, they seem to be saying that the agent's actual sequence of intentional states prior to the decision about how to act *could be continued* by forming the intention to take the better option, or (alternatively phrased) that the actual-sequence could turn out to be an earlier segment of an *R**-world. But this would be to say that the better option is *actually accessible* to the agent (or within the power of his mechanism *M* to bring about) at just the point where he (or *M*) decided to take the worse option.

Without fully realizing it, Fischer and Ravizza are pulled towards such a libertarian formulation – even though it is inconsistent with semicompatibilism – because *only* such an account of the freedom-aspect (a) of *akrasia* in terms of "regulative control" over better and worse options can also do justice to the weakness-aspect (b). The evidence for this is that all the obvious alternative accounts in terms of mere "guidance control" fail to save the (b)-aspect of weak-will phenomena. For suppose the semi-compatibilist says that the agent's practically irrationality can

[65] Fischer and Ravizza, *Responsibility and Control*, p. 73.

[66] As Fischer and Ravizza rightly insist, actual-sequence mechanisms can have dispositional properties defined in terms of the way the mechanism would function in possible worlds that are nevertheless not accessible to the agent, i.e., worlds in which he would do *X*, despite the fact that he cannot actually bring it about that he do *X* (Fischer and Ravizza, *Responsibility and Control*, p. 53). But the problem of weak-willed psychological mechanisms cannot be solved this way.

be understood in this way: the mechanism on which she acted (*M*) had a disposition to react to the sufficient reason to do otherwise (*R*), as recognized by the agent in the actual sequence, but did not. Then we need an explanation of *why M* did not function in the actual sequence as it was *disposed* to do, and thus produce in the agent the intention to act otherwise for the reason actually recognized.

There seem to be three ways that the semi-compatibilist could try to explain this failure of *M* to react to *R* as it was disposed to. First, perhaps *M* began to react to *R* and was blocked from forming the alternative intention by some external force. Then this is not a case of true *akrasia*, because the external element, rather than the agent, will stand (in the intentional order of explanation) directly behind the actualizing of the worse option. Second, a random malfunction in *M* could occur; but this also fails as a replacement-analysis for "agent-weakness." If a psychological mechanism is "disposed" to act on actually recognized sufficient reasons to do otherwise, but *just by chance* does not respond to them, then it is not a weak-willed mechanism, but simply an unlucky one. The practical irrationality of the resulting act is attributable to chance, rather than to the agent. Third, perhaps what intervenes to prevent *M* from acting on *R*, as it "can" do in the dispositional sense (or as its disposition *D* would indicate), was *another* mechanism of the agent. But this sort of interference will also reduce ultimately to bad luck (consider, for instance, ancient Greek conceptions of *akrasia* as the product of conflicts between practical reason and other psychic states of appetite and anger). Alternatively, perhaps the intervening mechanism is one whose intentional process aims precisely at causing the failure of more rationally disposed mechanisms to operate properly. If this is intelligible,[67] it would still need to be made clear how such a "spoiler-mechanism" could itself count as moderately reasons-responsive. What if it too is only weakly reasons-reactive (and hence capable of weakness or failure to perform its function in causing weakness or failure in other mechanisms)?: this would seem to generate a vicious regress. But even if this could be avoided, explaining weakness of will by the intervention of a spoiler-mechanism would seem to convert it into something closer to self-deception, rather than the true agent-weakness or the obstinate perversity we know all-too-well by (first- and third-personal) acquaintance.

[67] Note that to make this intelligible, we probably need to move away from a merely dispositional characterization of psychological mechanisms, and towards a more fully teleological characterization of them in terms of their proper functioning, or their designed or natural or optimal outcomes. This might give a Stump-style semi-compatibilist more chance of meeting these difficulties (but I suspect in the end, even a more Thomistic semi-compatilism will not meet the challenges of weakness of will and radical evil).

These three explanations seem so unsatisfying because what we want to say, in light of our experience, is rather that the agent (or her mechanism) *had the power* to react to recognized reasons for doing better than she did, but that she actively omitted *exercising* this power. This is more than saying that it was simply "possible" in the dispositional sense for the agent to do better. In actual-sequence models of *akrasia*, nothing can fill this role of exercising or not exercising the power to actualize the better option. Instead, on these models, the possibility of the better action merely *exists* for the mechanism on which the agent acts, and it is either realized or not, but beyond these brute facts there is nothing more to the story. Yet our intuitions tell us that there often is something more in the real phenomena, and this missing element is precisely what we mean by agent-weakness. Thus weakness of will is bound to be a stumbling block for actual-sequence models of moral responsibility. For either weakness is reduced to arbitrariness or bad luck, or the *agent* directly weakens her will by an exercise of her libertarian freedom: she perversely chooses to ignore the sufficient reason she recognizes for acting in a better way. This will of course seem mysterious, but perhaps only because it reveals something otherwise unapparent about human freedom. If so, then the right account *should* be mysterious in this way, since such perversity *is* as weird as it is real.[68]

The actual-sequence model indeed eliminates the mystery of volitional perversity, but precisely for this reason the actual-sequence model seems wrong. The actual-sequence alternatives are not mysterious in the right way, i.e., in the way that genuine weakness of will requires. For it would not be mysterious, but rather *nonsensical*, to say that the psychological mechanism on which the *akrates* acted was disposed to perversity, or disposed not to react to the actually present reason to do otherwise, even though it could react to it (understood as meaning that it *was* disposed to react to it). Yet this is what Fischer and Ravizza must say: the agent's mechanism has two opposing dispositions and one simply wins out. Their only other alternatives amount to offering us some phenomenologically inadequate substitute for agent-weakness, such as a properly functioning mechanism that would have acted rightly but for the interference of some outside force or some random chance. Thus the actual-sequence account avoids mystery only by misrepresenting reality,[69] i.e., denying that there is

[68] This is why Søren Kierkegaard devoted an entire book, *The Concept of Anxiety*, to analyzing this perversity in some of its many forms.

[69] Analogously, note how much less mysterious it would have been for us if the microphysical universe did not work according to irreducible quantum probabilities, but respected nice, rigid, picturable Newtonian processes.

genuine agent-perversity, in which the agent (or her mechanisms) directly brings about her volitional weakness when she could have avoided bring it about. Here the convenient lack of mystery offered by semi-compatibilist theory is suspect: we cannot honestly deny the existence of true agent-perversity, since we find it in ourselves.

4. CONCLUSION

Fischer and Ravizza's new account of moderate reasons-responsiveness is meant to be a philosophical reconstruction of the concept of moral sanity. Given the woefully inadequate understanding of sanity as a condition of mroal responsibility in our legal tradition, we badly need the kind of enlightenment that Fischer and Ravizza's theory can provide. But as the authors are well aware,[70] sanity is not an exact concept; there are going to be many borderline cases. Philosophers cannot be expected to go into too much detail in specifying the sorts of sensitivity to reality, responsiveness to others and to community, stability of psychological mechanisms over time, and cognitive coherence in processing reasons required for sanity. At some point this becomes the work of psychological theory. Fischer and Ravizza's real contribution in this area is to suggest that, however psychologists may flesh out the contours of this complex phenomenon, the basic conditions of sanity can be explained in terms of dispositional properties of the agent's mind, which *do not* require that the agent really be able to think or act otherwise than he did (or be able to bring about that he have alternative preferences or form alternative intentions, etc.) in the same initial state. Sanity is thus disconnected from libertarian freedom or "regulative control." This approach is promising in many respects although, as I have suggested, the phenomena of sane but weak-willed agents may be impossible for it to handle. If so, then libertarian freedom may not be dispensable after all.

Department of Philosophy
Fordham University
Lincoln Center
113 West 60th Street
New York, NY 10023, USA
E-mail: davenport@fordham.edu

[70] Fischer and Ravizza, *Responsibility and Control*, p. 80.