Many philosophers are convinced by Harry Frankfurt’s (1969) controversial argument that moral responsibility does not require the ability to do otherwise. According to Frankfurt, it is possible to construct an example in which an agent acts freely and is morally responsible for what she does but is unable to do otherwise. Frankfurt’s example featured Jones, who shot Smith for his own reasons. As it happened, Black wanted Jones to shoot Smith on his own, but covertly arranged things so that if Jones were about to fail to do so, Black would manipulate Jones so that he (Jones) would shoot Smith. Since Jones shot Smith on his own, Black remained dormant; he played no role in Jones’s act. Jones, it seems, acted freely and was morally responsible for doing so. Yet because of Black’s presence, Jones could not have done otherwise. Hence, as the example involving Jones illustrates, moral responsibility does not require the ability to do otherwise.

1 This paper was written for the 2011 New Orleans Workshop on Agency and Responsibility (NOWAR), organized by David Shoemaker. I would like to thank him for arranging this excellent event. I delivered an earlier version of this paper to the University of Texas at Austin Philosophy Department. I am indebted to various members of the audience for their instructive comments and criticisms. Notable among them were Sinan Dogramaci, Jeremy Evans, Alex Grzankowski, Bob Kane, Chris Simpson, Nicole Smith, David Sosa, and Michael Tye. At NOWAR, I profited from discussion with many in attendance, especially Michael Bratman, David Brink, Sarah Buss, Stephen Kearns, Rahul Kumar, Dana Nelkin, Paul Russell, Tim Scanlon, Josh Sheppard, David Shoemaker, (yet again) Nicole Smith, Jada Strabbing, and Gary Watson. For further helpful advice, I would also like to thank John Martin Fischer, Ish Haji, Terry Horgan, Al Mele, Shaun Nichols, Derk Pereboom, Carolina Sartorio, and Brandon Warmke. Finally, I would like to thank two anonymous referees for Oxford University Press.

2 For an examination of the further details offered in support of Frankfurt’s argument, see the introduction to Widerker and McKenna (2003), and also my (2008). Among those who defend some variation of Frankfurt’s argument are Fischer (1994),
Granting that moral responsibility requires some kind of freedom, the lesson that those convinced by Frankfurt’s argument should take from it is that the kind of freedom exercised by Jones, whatever it comes to, is not to be accounted for in terms of the ability to do otherwise. Hence, it is not to be understood in terms of alternatives to the way an agent acts. It is, instead, to be understood in terms of the source of her actions. It requires, as John Martin Fischer (1994) has put it, attention to the actual-sequence of events leading to action. Call this kind of freedom source freedom, and the kind that is understood in terms of alternatives leeway freedom.3

Despite what many assume, Frankfurt’s conclusion is not the exclusive domain of compatibilists. Source incompatibilists have accepted the conclusion to Frankfurt’s argument and set out to develop an actual-sequence account of free action, one that requires the falsity of determinism.4 Nevertheless, compatibilists stand to gain more should Frankfurt’s argument turn out to be sound. This is because incompatibilists have a powerful argument for the conclusion that determinism is incompatible with the ability to do otherwise. This argument, the consequence argument, strongly suggests that if free will and moral responsibility require the ability to do otherwise, then determinism is incompatible with both.5 Armed with Frankfurt’s argument, source compatibilists can grant that determinism is incompatible with the ability to do otherwise. But then they can argue that, even if it is, this is not relevant to whether moral responsibility and (all of) the freedom required for it are compatible with determinism. As a source compatibilist, this is how I proceed.6

The project for the source compatibilist who relies upon the soundness of Frankfurt’s argument is to offer a plausible actual-sequence account of source freedom that satisfies two desiderata. First, it is of a sort that can be


3 I write in terms of “source freedom,” while Fischer (1994), and Fischer and Ravizza (1998) use the term “guidance control.” I regard these terms as amounting to the same thing. Likewise, I prefer the expression “freedom to do otherwise” or “leeway freedom,” whereas Fischer and Ravizza write in terms of “regulative control.”
5 The consequence argument was first developed by Ginet (1966), and then subsequently refined by van Inwagen (1975), and Wiggins (1973).
6 The person who deserves the credit for this way of framing the dialectic is Fischer (1994). Fischer coins a very similar view “semicompatibilism.” Where I differ from him is that, on the basis of (some version of) the consequence argument, I am prepared to grant to the incompatibilist the incompatibility of the ability to do otherwise. Fischer instead wishes to remain uncommitted on this point.
present in a Frankfurt example; second, it can be shown to be compatible with determinism. Although many source compatibilists are attracted to a mesh or hierarchical theory of the sort that Frankfurt himself endorses, in my estimation, such theories face insurmountable problems. In what follows, I shall instead explore the prospects of accounting for source freedom in terms of responsiveness to reasons. In so doing, I will pay special attention to a distinctive approach thoughtfully developed by Fischer and Mark Ravizza (1998) which focuses on the mechanisms of an agent’s actions. As I shall explain, a mechanism-based approach appears to be necessary in order to fit a reasons-responsive theory for Frankfurt examples. Unfortunately, this approach comes at a high cost, since there are serious problems with analyzing exercises of source freedom in terms of an agent’s mechanisms of action. Ultimately, I shall seek an account of reasons-responsiveness that avoids these difficulties.

1. THE APPEAL OF A REASONS-RESPONSIVE THEORY

Reasons-responsive theories have an ancient lineage that arguably can be found in Aristotle. They are appealing irrespective of the compatibilism-incompatibilism debate for at least two reasons. First, they target a feature of agency distinctive of persons, a feature that marks persons as rational animals. Second, they offer elegant explanations of the conditions in which various excuses and exemptions apply—the persons in question, it is contended, are not suitably responsive to a proper range of reasons. For instance, a plausible way to distinguish between the person who, weak willed, freely drinks the beer from the person who is compelled by her addiction to do so is that the former is responsive to a wider range of reasons for not drinking than is the latter.

Compatibilists have found powerful reasons to develop a reasons-responsive theory quite apart from the unique dialectical burdens of source compatibilism. Reasons-responsiveness seems especially well suited for a compatibilist analysis since in explaining sensitivity to reasons, we must consider how an agent would respond to a range of reasons and not just those present to an agent when she performs an action in the actual circumstances that she is in. This fits naturally with attending to a range of dispositional or modal properties of agency, and there is little reason to

---

7 See Appendix I for my source compatibilist objection to these theories.
8 See Terrence Irwin’s (1980).
think that determinism is incompatible with the possession of these sorts of properties generally.

To appreciate how a reasons-responsive theory might be developed, begin with the following two preliminary points:

First, when an agent acts who is suitably reasons-responsive, the most important factor for the source compatibilist in accounting for her freedom is that the etiology of the act which she actually performed involved springs that were sensitive to reasons. For now, let us think of those springs in terms of the agent herself as the cause of her acts, where this is not meant to commit to thinking of agents as distinct, irreducible substance causes. Different reasons, understood as different inputs, would have yielded different outputs, understood as alterations in modes of conduct. And what this shows is that the agent’s response to the actual “inputs” played a role that was itself sensitive to, or responsive to, the actual conditions in which the agent acted. To illustrate, consider a simple example of the sensitivity of a primitive gizmo, a thermostat. Suppose a thermostat is set at 76 degrees (Fahrenheit) and the room the thermostat is in turns out to be 76 degrees. One might wonder if the thermostat’s setting accounts for the temperature of the room. After all, it might be disconnected and so merely a fluke that its setting and the room’s temperature are 76 degrees. When we learn that the room would come up to 78 were the thermostat set to 78, or would come down to 74 were the thermostat set to 74, and so on for numerous other values high and low of 76, we do not just learn something about the way the thermostat would behave; we also learn about how, in the actual scenario when it is set to 76, it does behave; it plays a certain causal role from reliable and suitably sensitive resources.

The same can be claimed about a person who is responsive to reasons. Consider by contrast a clear case of reasons-responsive failure, a case in which a person fails to take reasons as inputs in a manner that yields proper results (like a defunct thermostat). Imagine a compulsive hand-washer washing her hands after handling some trash. Not knowing her condition, we might initially think that her freedom consists in part in her responding rationally to conditions that would warrant hand washing. But now suppose that we learn that she’ll wash her hands when they brush up against a bottle of bleach, or when she touches her own shirt, or if the sun shines on them, or if the wind blows, or if Don Knots is on TV, or there is a Christmas carol playing on the radio. This casts doubt on whether the reason she is washing her hands in the actual world is because she was handling some trash. Or, even if her washing her hands is to be accounted for in terms of her handling some trash, her insensitivity to other reasons suggests that the manner in which she caused her action was not suitably
sensitive to the significance of trash-handling, as in contrast with, say, the wind blowing.

Second, responsiveness to reasons comes in degrees. This raises questions about what amounts to an adequate degree of responsiveness and what sorts of failures are tolerable. Consider again the thermostat. Suppose the thermostat functioned as described for uses in a normal home environment. It would not impugn the adequacy of the thermostat to learn that it is not super-responsive, and so not responsive to temperatures of 2 degrees at one end and 150 at the other. Alternatively, the thermostat might be only very minimally responsive, functioning properly between settings of 75 through 77, but yielding whacky results at any other settings. Furthermore, failures of responsiveness can be localized. Imagine that the thermostat that is set at 76 would not have responded differently to any settings of between 74 and 78, but would have responded properly to settings of lower than 74 or higher than 78. Here, we could still not determine what role setting the thermostat at 76 played in the room’s being 76 degrees, since the room would have remained at 76 degrees with settings of as low as 74 or as high as 78. Even if we were to conclude that setting the thermostat at 76 caused the room to become 76 degrees, we would be right to think that the mode of its causing the change in temperature was too fluky, so much so that all we could feel confident in claiming as a credible explanation is that it reliably caused the room’s temperature to be some temperature anywhere between 74 and 78, and as chance would have it, it happened to be 76.

The same points apply to persons and their responsiveness to reasons. A person who is suitably reasons-responsive in a way that is sufficient for free action need not be able to respond to a range of reasons that meets exceedingly high standards, otherwise almost no one would turn out to act freely. And a person who is only responsive to a very slim range of reasons will fall short of acting freely while nevertheless being responsive to some reasons. Take our compulsive hand-washer. It seems that she will wash her hands come what may. But it might be that a very limited range of incentives would result in her not washing her hands: if her child were on fire, or if her favorite Mozart concerto were playing and she instead wanted to dance, and so on. While such cases might show her to be responsive to reasons, she would only be minimally so, and this would not be enough to regard her as acting freely in washing her hands as she did. Furthermore, as with simple gizmos like thermostats, so too with persons, localized failures are possible. Someone who is as competent as anyone might fall to pieces in the presence of an abusive spouse or a domineering partner. Localized foibles are part of the package deal for most of us imperfect humans.
Now, in light of the two preceding points, let us examine Fischer and Ravizza’s carefully developed reasons-responsive theory. According to them, at one end of the spectrum is strong reasons-responsiveness (SRR). An agent who is SRR with respect to an act is such that, if there were sufficient reason for her to do otherwise, she would do otherwise (Fischer and Ravizza, 1998: 41). At the other end of the spectrum is weak reasons-responsiveness (WRR). An agent who is weakly reasons-responsive with respect to an act is such that there is at least one (nonactual) scenario in which there is sufficient reason to do otherwise and the agent does otherwise (44). As should be clear, and in keeping with the second point developed above (in Section 1), SRR is too strong; if a condition of free agency required SRR, it would turn out that almost no one acts freely and is morally responsible for what she does. On the other hand, WRR is too weak; if a condition of free agency required no more than WRR, it would turn out that many severely impaired agents would fully satisfy the freedom conditions for moral responsibility. Our compulsive hand-washer was WRR in washing her hands since she’d have not washed them if her children were on fire. So, what is needed, and what Fischer and Ravizza develop in admirable detail, is a middle-of-the-spectrum degree of responsiveness, moderate reasons-responsiveness (MRR), that is suited for sane, competent, albeit imperfect, moral agency—the sort of agency of which most psychologically healthy adults are capable while functioning in normal practical contexts of deliberation and action (69–84).

So how do Fischer and Ravizza develop MRR? Understanding their proposal requires careful attention to what is involved in being responsive to reasons. Fischer and Ravizza distinguish between a receptivity component and a reactivity component (1998: 41). Being reasons-receptive is a matter of being able to recognize what reasons there are, and in particular, being able to recognize what reasons count as sufficient for action. Being reasons-reactive is a matter of being able to react to the reasons one recognizes as sufficient by choosing and acting as needed. According to Fischer and Ravizza, MRR requires moderate reasons-receptivity. For a person to be morally responsible for what she does, she must be receptive to reasons in

---

9 In examining Fischer and Ravizza’s view in this section, I shall depart from their doing so in terms of mechanisms that are reasons-responsive and that are owned by the agent. I shall instead just write in terms of agents being reasons-responsive. In the next section, I’ll recalibrate to fit their mechanism-based formulation.
a manner that displays a rich pattern of recognition whereby reasons can be placed on a scale with a continuum involving stronger and weaker incentives. For instance, if an agent regarded an incentive of $100 as sufficient for acting in a certain way, she would take higher values as also sufficient, while being open to taking lower values as insufficient. Furthermore, the agent’s pattern of reasons recognition must display a sane appreciation for reasons that are grounded in reality, and the agent must also be able to recognize moral reasons and be able to grasp that sometimes they are sufficient for acting in ways that morality prescribes. All of this, of course, is consistent with being an agent who is not receptive to the full spectrum of reasons for how she ought to act, including the full spectrum of moral reasons.

As regards reasons-reactivity, Fischer and Ravizza allow for a striking asymmetry. An agent who is MRR need only be weakly reasons-reactive. It is enough, they contend, that among the worlds in which an MRR agent is receptive to sufficient reasons to do otherwise, she react to only one of those reasons. Here, one might wonder why weak reactivity would be enough. Wouldn’t this show that in the wide spectrum of cases in which the agent recognizes sufficient reasons but does not act upon them that the agent cannot act upon them and so is impaired for morally responsible agency? Fischer and Ravizza do not think so. They contend that “reactivity is all of a piece” (1998: 73). As they see it, the fact that there exists just one world in which an agent reacts differently to a sufficient reason to do otherwise is sufficient to establish that in each world in which an agent recognizes sufficient reasons to do otherwise that she is able to react to those reasons, even if at those worlds she does not.

Critics have pressed Fischer and Ravizza on various details of their proposal regarding the spectrum of MRR. One has to do with their requirement of weak reasons-reactivity. As I have argued elsewhere (2005), weak reactivity looks to be too weak. The fact that an agent would react to only one reason to do otherwise but not a constellation of other similarly related reasons calls into question her agency. Hence, I have proposed a variation on MRR in terms of weaker reactivity, not weak reactivity. The spectrum of reasons to which an agent must be reactive can be weaker than the spectrum of reasons to which she is receptive, but

---

10 Fischer (2006: 328) has granted this point in response to Mele (2006b: 290), who presses a similar worry.

11 This is needed in order to allow for cases in which an agent freely and knowingly acts contrary to what she has sufficient reason to do. Note, however, that I have only claimed here that the spectrum of reactivity can be weaker than that of receptivity. Elsewhere (2005), I have made the stronger claim—that their view requires this asymmetry. Dana Nelkin and David Brink have convinced me that this is a mistake. The requirements of morally responsible agency do not, strictly speaking, require any asymmetry. Nevertheless I suspect for nearly all actual persons who are morally responsible
the spectrum of reactivity still needs to display a sane, stable pattern along the lines of moderate receptivity. There are further details that we might add here to supplement Fischer and Ravizza’s proposal, or instead various details of their thesis with which we might take issue, but for present purposes the preceding discussion will do.

While it is important to make clear the delicate effort Fischer and Ravizza make to get right the degree of responsiveness needed to account for MRR, the most important factor in their account of freedom is that the etiology of an agent’s act involved springs that were themselves sensitive to reasons. This shows Fischer and Ravizza’s deference to the first of the two preliminary points set out above (in Section 1). I pause here to emphasize this since it is easy to lose sight of. The salient point to note is that by demonstrating that when an agent acted she was MRR, one does not just show how that agent would respond to other reasons if those reasons were present to her. One also shows that in acting as she did, the agent was responsive to the conditions she was actually in. Her actually-operative springs of action were functioning in ways that themselves were responses to the conditions in which she actually found herself.

To elaborate on the preceding point: It is easy when reflecting upon Fischer and Ravizza’s proposal to misunderstand the theoretical purpose to which certain counterfactuals are put. Consider a MRR agent who in deciding to remain home and work on an article would be receptive and reactive to the reason that the legendary blues guitarist Buddy Guy is putting on a concert that night and she could easily go. Were she presented with this reason, she would not work on her article. Instead, she’d go to the show. Don’t be snookered here into thinking that what this counterfactual is meant to establish is that, in acting as the agent did, she could have done otherwise. That is, the point of this counterfactual, as well as other counterfactuals involving other reasons to which this agent might be both receptive and reactive, is not meant to establish that the agent had a freedom that concerned alternatives to what she did—leeway freedom. It is instead meant to establish that in acting as she did, she was sensitive to reasons in such a way that her actual bringing about of her act of remaining home and working on her article was itself suitably sensitive to rational considerations. Some compatibilists, such as those classical conditionalists like Ayer (1954), Davidson (1973), Hobart (1934), and Moore (1912), attempted to rely upon similar counterfactuals in order to account for a compatibilist theory of the freedom to do otherwise. But Fischer and Ravizza do not agents, there will be some asymmetry of this sort; most persons are not moved to act by the full range of sufficient reasons for action to which they are receptive.
understand the worlds in which an agent does otherwise in response to differing reasons as worlds that, as they put it, are accessible to the agent in the actual world in which she does act (1998: 53). So, as they see it, it cannot be that by virtue of such worlds an agent in the actual world is able to do otherwise. Hence, Fischer and Ravizza do not mean for these counterfactuals to aid in underwriting an account of the freedom to do otherwise.

3. SHIFTING TO A MECHANISM-BASED THEORY

Given the preceding treatment, can we fit a reasons-responsive theory into the actual-sequence constraints of Frankfurt’s argument? Regrettably, on first appearance, these two compatibilist-friendly theses do not appear to make comfortable bedfellows. Recall Jones from the Frankfurt example presented earlier. Jones, says the source compatibilist, acted freely. But due to Black’s presence, whatever reasons might have been put to Jones, it is not the case that he would have done other than as he did. This is because Black was playing the role of a “counterfactual intervener” and stood prepared to ensure that Jones would shoot Smith no matter what. Thus, if an agent acts freely in a Frankfurt example, it seems that her freedom cannot be explained in terms of her responsiveness to reasons. Due to the presence of a counterfactual intervener, the agent would not act otherwise even if different reasons were present (1998: 38). So is reasons-responsiveness ill-suited for a source compatibilist theory?

According to Fischer and Ravizza, we can fit a reasons-responsive theory to the kind of freedom present in a Frankfurt example by shifting from an agent-based to a mechanism-based theory. The agent in a Frankfurt example, they contend, is not responsive to reasons, but the mechanism from which she acts, which is her own mechanism of action, is reasons-responsive. To appreciate their intriguing proposal we need to answer two questions. First, what is a mechanism of action as they understand it? Second, how is it that focus upon an agent’s own mechanism of action ensures immunity to the problem that seems to arise for an agent’s reasons-responsiveness in a Frankfurt example? Why does the shift to mechanisms help?

As regards the first question, the mechanism of action, as Fischer and Ravizza understand it, is meant to pick out “nothing over and above the process that leads to the relevant upshot” (1998: 38). To make clear that they do not mean to reify the notion of mechanism or to suggest that it is something like a natural kind, they remark that rather than use the term, we could instead just speak in terms of the “way the action came about” (38).
Clearly, in this sense, the mechanism of action at work in the Frankfurt example involving Jones’s acting on his own is very different from the mechanism that would have operated were Black to have intervened and caused Jones to shoot Smith. That much, it can be safely granted, is clear and uncontroversial. Furthermore, one can also infer that among the full constellation of events, states and processes constituting Jones, or Jones’s complete psychic condition at the time, not all of the elements of that constellation were causally implicated in Jones’s shooting Smith. Jones, for instance, might have a deep love of his mother, a fondness for kittens, and be especially skilled at solving calculus problems. But these, we can safely assume, can be “filtered out” of the complex of Jones’s psychic events, states, and processes that were implicated in the process leading to his act of shooting Smith. Thus, an agent’s mechanism of action, however it is to be construed, carves out a subset of the full set of psychological elements constituting the agent’s psychology.

As regards the second question, Fischer and Ravizza argue that to test whether in a Frankfurt example an agent’s own actually-operative mechanism of action was reasons-responsive, we need to test it for sensitivity to various reasons in conditions in which it operates uninhibited, and to do this we have a license to “go to worlds”, as the expression goes, in which Black’s presence is subtracted. So, were Black not present, and were Jones presented with pertinent reasons for not shooting Smith, Jones, by way of his mechanism of action, would react otherwise and not shoot Smith. According to Fischer and Ravizza, in order for it to be the case that Jones exercises adequate source freedom in shooting Smith, his mechanism of action must be such that the pattern of reasons to which it is sensitive would be MRR.

4. PROBLEMS WITH MECHANISMS

Various critics have objected to Fischer and Ravizza’s mechanism-based approach. R. Jay Wallace has remarked that by shifting from an agent-based to a mechanism-based view, the “intuitive locus of responsibility, the person, drops out of view” (1997: 159). He claims that instead we should tend to the normative competence of persons. Gary Watson has expressed his concern that characterizing a mechanism of action in terms that are no more specific than “‘a process resulting in behavior’ is too amorphous to do the refined work needed by the theory” (2001 as appearing in Watson, 2004: 299). And Carl Ginet has claimed that the postulation of mechanisms is not doing any useful work in their theory (2006: 235–6).
According to him, Fischer and Ravizza want it to be that an agent would respond differently to certain reasons by virtue of a mechanism. Their theory, Ginet contends, could be preserved by cutting out the middle term and just theorizing in terms of an agent being reasons-responsive. I too have registered some concerns with their appeal to mechanisms. However, my criticisms have been more limited. I have only argued that Fischer and Ravizza need to provide further support for their account of mechanisms in order to avoid various objections.

In my estimation (2001), Fischer and Ravizza face difficulty with individuating the mechanisms operative in exercises of free action. They resist pressure to provide any principled basis for mechanism individuation and contend that they can make do with an intuitive sense of what counts as the same mechanism and when, in shifting between contexts, we move to different mechanisms (1998: 40). Clearly, there is a different mechanism of action at work when an intervener such as Black causes Jones to act as in contrast to when Jones acts on his own. Thus, on their view, just as we are able to reidentify the same face or smile, or the same house or car, without a particular theory or general principle, so too we can recognize same mechanism of action operative in different contexts. The problem, however, is that there are some contexts crucial to the development of their theory where, absent some principled bases for mechanism individuation and differentiation, it is hard to assess their broader claims. Obviously they are in no danger when distinguishing between actual and alternative mechanisms in Frankfurt examples. But it is an indispensible part of their theory that we are to test how the same mechanism behaves in response to different reasons. How are we to settle whether, when different reasons are put to an agent, the same mechanism is operative? We need some purchase on what it is that we are holding fixed when we hold fixed “the same mechanism” while testing its degree of responsiveness.

The problem stems from Fischer and Ravizza’s lean characterization of a mechanism of action. “The process that leads to action” or “the way the action came about” allows for further specification by way of extremely narrow or instead extremely wide descriptions. Picking out the process that

---

13 They do offer one theoretical constraint. The descriptive content by which the mechanism is identified must be “temporally intrinsic” (1998: 46–7). That is, it must pick out the mechanism in such a way that it does not involve reference to later times. Otherwise, the mechanism would require certain outcomes, and thus would not permit variability in the face of different reasons. Hence, the temporally extrinsic mechanism “deliberation prior to donating to a charity” is ruled out since it entails acting in a certain way.
leads to action seems most naturally to involve the entire complex of proximal antecedent states and events figuring into a correct explanation of the pertinent action.\textsuperscript{14} Suppose that we limit the items in this complex just to the agent-involving ones—an obviously permissible restriction. Still, is every agent-involving feature of the complex to be included in the description? What salient features might one chose to whittle the description down? At one level of description, sameness of mechanism right down to microdetails would be required. But, as I have argued elsewhere (2001: 97), this would likely give rise to serious problems. Suppose that the blameworthy reason upon which an agent acts in the actual world—or rather the set of psychological ingredients involved in recognition of this reason—is identical with or supervenes upon some neurological state of the brain. If sameness of mechanism requires sameness of microdetails, when \textit{that} mechanism is tested to learn how it would respond to different reasons, it would have to respond to different reason by resources that shared the same microdetails with the brain states involved in recognition of the blameworthy reason in the actual world. This is nomically unlikely, albeit, admittedly not metaphysically impossible. It thus exposes Fischer and Ravizza’s theory to the rather likely prospect of empirical refutation. So it seems that a very narrow description needs to be ruled out.

Of course, Fischer and Ravizza would be on solid ground in ruling out such narrow descriptions.\textsuperscript{15} Typically, when identifying causal processes or ways events are brought about, we do not mean to focus upon microdetails. For instance, identifying the event of the flood as the cause of the erosion needn’t require that we focus on the microdetails of just how it was that the water molecules were arranged. In such contexts it would be highly unreasonable to demand such a level of specificity. Fischer and Ravizza are certainly entitled to a similar reply here as regards the sorts of processes leading to (putatively) free acts.

But then, how wide, and so permissive, should the needed description be? Because Fischer and Ravizza offer no principled basis for mechanism individuation, we are left to settle the matter exclusively by appeal to our intuitive reactions to varying cases. Unfortunately, intuitions in these kinds of highly theoretically charged contexts can vary widely. The neurophysiologists’ modes of parsimony, the cognitive scientists’, or instead the cognitive therapists’, are going to be very different from those employed in normal folk-psychological discourse in which typical claims

\textsuperscript{14} I am indebted to Carl Ginet for this point. See also his discussion of it (2006: 233).

\textsuperscript{15} I am indebted to Derk Pereboom, who emphasized this point.
of blameworthiness and praiseworthiness have their natural homes. And among philosophers differentially committed as regards the free will dispute, it is easy to see that some will be inclined to a hyperrestricted description that will narrow in on requiring same past and laws. Whereas others will be open to descriptions that are far more permissive. Of course, this would just be to redraw old lines in new sand. So that won’t help. What is needed, it seems, is some rationale with theoretically independent appeal that will allow us to assess claims of same or different mechanism.16

5. A DEEP PROBLEM FOR ANY MECHANISM-BASED APPROACH

The preceding considerations strongly suggest that if a mechanism-based version of a reasons-responsive theory is to survive, there must be some principled way to get an independent purchase on what counts as the same mechanism across relatively diverse contexts. In my earlier work (e.g. 2001, 2011), I have often expressed confidence that there is a way to supplement Fischer and Ravizza’s theory so as to offer the right kind of rationale for mechanism individuation. Unfortunately, my considered opinion now is that there is a structural problem which stands in the way of offering any more specific content to the individuation conditions for mechanisms of action. To explain, it will be useful to begin by considering an objection developed by Ginet (2006: 234–5).17 Fischer and Ravizza take up the example of a person driving a car from a mechanism of unreflective habit (1998: 86). This person takes an exit she often takes, thinking nothing of it. Were it blocked, she’d respond properly to the pertinent reason without any deliberation at all. She’d just head to the next exit. But as Ginet plainly points out, while the case Fischer and Ravizza consider is readily handled by their treatment, there is a slight variation on it which is problematic. Imagine instead a person who, upon seeing the exit blocked, is unsure of what other exit she ought to use. Here, what would be most natural would be for her to deliberate, and it would seem that a reasons-responsive theory should seek to accommodate such a case. But Fischer and Ravizza hold fixed the mechanism of unreflective habit. Hence, they cannot permit that

16 For an illustration of one particular point in which, in responding to an incompatibilist concern, Fischer and Ravizza cannot make do without a principled basis for mechanism individuation, see Appendix II. There, I also discuss Fischer’s (2004: 166–71) thoughtful reply to my criticism.

17 Watson has raised a similar worry (2001 as appearing in 2004: 298).
the mechanism of unreflective habit from which this agent acts is plastic enough to allow reflection when conditions call for aborting reliance on habit alone.

Ginet’s insightful criticism is a specific instance of a more general problem that is bound to plague any attempt to give more content to what might count as the conditions for individuating mechanisms. The general problem, as I see it, is that any complex system will have “subsystems” that are designed to function precisely by shutting down or by permitting other systems to override in some contexts but not others. Assuming a person functioning as a practical agent can be understood as a complex system, or at least as relevantly analogous to one, any attempt to whittle down which agential elements are the ones implicated in “the” mechanism of her action is bound to restrict the agent’s flexibility as regards reasons-responsiveness.

The postulation of an agent’s mechanism of action as narrower than the full person herself looks innocent enough when one is invited to consider a case like Jones and note that Jones’s ability to do calculus is irrelevant to the agent-involving factors causally contributing to the shooting of Smith. There are likely lots of these ingredients in Jones’s complete psychic constitution that could be whittled away: his love of his mother, his fondness for kittens, and so on. But once we seek to put more meat on a restriction of a pertinent mechanism of action, we will face problems like the one Ginet has identified. If we back off and allow for the description of the mechanism to be “fatter” so as to allow for shifts between processes like unreflective and then reflective habit, we pretty much have what is the functional equivalent of the agent herself as the mechanism, minus various outlier traits like the ability to do calculus.

Reflection on complex mechanisms like automobiles or computers, things that are built up out of smaller mechanisms, helps to illustrate the point. Suppose that one of a computer’s programs runs unimpeded, but the computer is designed to divert that program and prioritize other operations if the system is under stress, or uploading new software, or about to lose power, or what have you. The same applies to an automobile. A car will allow things like unimpeded gas flow through the fuel injection system unless another part of the system recognizes problems with the fuel mix or something of the sort. If we were to evaluate the degree of sensitivity of such a system by “holding fixed the actually operative sub-mechanism” we’d hamstring the system for a variety of conditions to which, without holding these fixed, the larger system as a whole would be able to respond quite easily.18

18 An anonymous referee for OUP has astutely asked about the following way of defending Fischer and Ravizza’s appeal to mechanisms. To illustrate, consider Ginet’s example of the driver acting from the mechanism of unreflective habit. Why is it a
To give just one action-theoretic example of this sort of thing, consider Alfred Mele’s (1995, 2006) work on free agency (sometimes he writes in terms of autonomy). One element in Mele’s approach is his account of both weakness and strength of will. To account for these phenomena, Mele distinguishes between the processes by which an agent judges what it is best to do from the motivational ingredients involved in directing her actions. Often, when a person acts, her evaluations of the objects of her desires are aligned with the strength of the desires themselves. In these cases, when she acts, the resources by which she might be able to exercise strength of will and avoid temptation are, so to speak, off line. They’re causally inert, and so could not be counted as anything like part of the actual causal process issuing in action. But the agent might be such that she has a kind of monitoring system, and were it that conditions would present themselves (were certain reasons to arise) that would involve the agent desiring more something incompatible with what she judges it best to do, the agent would be able to rely upon these agential resources in such a way as to act with strength of will and thereby act contrary to what she most desires (Mele, 1995: 27). Mele’s model here is a perfect example of the kind of dynamic relation between “subsystems” within the overall complex of an agent as a whole. If we want to include in “the mechanism of action” this entire complex, we will have to open the descriptions for “same mechanism” to include more than just what is, narrowly construed, causally implicated in the process leading to action. Here, it seems, “the mechanism” will become so inclusive that we might as well simply identify the entire person with the mechanism.19

problem for Fischer and Ravizza that in some scenarios the agent would have to “shift” to a different mechanism by beginning to deliberate? Isn’t it enough that the mechanism of unreflective habit is responsive enough to certain reasons to allow for it to, so to speak, shut down and allow another to come on line? Why isn’t this enough to say that the mechanism is suitably reasons-responsive? In response, I acknowledge that this is after all a thoughtful way to attempt to defend Fischer and Ravizza. But the upshot is to do so by claiming that the full spectrum of reasons to which an agent (by way of mechanisms or otherwise) ought to be responsive for (something like MRR) requires the postulation of a plurality of mechanisms. At this point, it appears that it is not a matter of holding fixed a mechanism, and now one wonders how close we are getting to just straightforwardly talking of the responsiveness of persons.

19 In correspondence, David Shoemaker has thoughtfully expressed some skepticism about this last point—that a more inclusive notion of a mechanism of action will be so inclusive that we might as well identify it with the person, or with the functional equivalent of the person. In conversation, Shaun Nichols has as well, citing the work by Fodor on the modularity of the mind. Here, I have no problem leaving it as an open question whether there is some richer notion of mechanisms that would do the work that needs to be done for a mechanism-based view like Fischer and Ravizza’s. But I strongly suspect that even if such a kind of mechanism were identified, and even if it turned out to
So, are we back to the incompatibility of a reasons-responsive theory and source compatibilism? According to Fischer and Ravizza, in order for an agent to be reasons-responsive, it must be that if different reasons were presented to the agent, then for at least some range of reasons, the agent would react otherwise. In a Frankfurt example, they contend, due to Black’s presence, the agent is not able to react otherwise.

This result—the incompatibility of reasons-responsiveness and source compatibilism—would come as welcome news to those contemporary compatibilists who reject Frankfurt’s argument and defend the traditional association between free action, moral responsibility, and the ability to do otherwise. Call these traditionalists leeway compatibilists. Amongst these leeway compatibilists Michael Fara (2008), Michael Smith (2003), and Kadri Vihvelin (2004) have each independently built upon recent developments regarding the nature of dispositions to account for the ability to do otherwise. In a Frankfurt example, each argues, an agent retains the ability to do otherwise. Randolph Clarke (2009) has labeled the thesis these philosophers advance the new dispositionalism. It is open to the new dispositionalists to argue that in a Frankfurt example, an agent is reasons-responsive, because, in being able to do otherwise, she is able to react to different reasons.

While there are differences in how they execute their arguments, the new dispositionalists are united in proposing that the free-will ability is somehow to be accounted for in terms of dispositions. Following the efforts of philosophers like David Lewis (1997) and C. B. Martin (1994), these leeway compatibilists all pay heed to accounts of dispositions that incorporate lessons learned from the way dispositional properties can be masked or finked. A disposition is masked when its manifestation is concealed in some way. A piece of salt placed in water remains soluble even if, when placed in water, it does not dissolve because it is encased in wax. A disposition is finked when, in just those conditions that would otherwise trigger manifestation, it is altered so as not to have the disposition. A glass vase sitting on a shelf undisturbed possesses the disposition of fragility even if, were it knocked over, a wizard would turn it to stone before striking the ground, rendering it not fragile.

be narrower than the full person qua agent, it would be much richer in content than anything like something restricted just to what is actually involved in the causal generation of action.
Cases of masking and finking show that dispositions cannot be analyzed in terms of simple conditionals such as, “if the vase were toppled, it would break” or “if the salt were placed in water, it would dissolve.” The vase mentioned above was fragile, but because it was finked, it would not break when toppled. The salt was soluble, but because it was masked, it would not dissolve if placed in water. In slightly different ways, each of the new dispositionalists propose more complex counterfactuals the truth of which confirms the possession of pertinent dispositions. Roughly, their strategy is of the following variety:

SC: If this vase were toppled, and if it retained during the relevant duration of time its intrinsic properties P₁-Pₙ, and if it were not interfered with in a way that would impede the causal efficacy of those intrinsic properties, then it would break.

Now, SC is no more than a crude illustration of what a carefully worked out formulation should look like, the sort that, for instance, both Fara and Vihvelin consider. But for present purposes it will suffice. It is by virtue of counterfactuals like SC that the new dispositionalists propose a more sophisticated treatment of dispositions. If abilities are then to be explained (Fara, 2008), or, more vigorously, fully analyzed (Vihvelin, 2004), in terms of dispositions, then counterfactuals like SC provide a template for an explication or analysis of the free-will ability. An especially important factor in any such proposal is that the deeper explanation of the possession of the dispositional properties by an object concerns the relevant intrinsic properties possessed by the object (e.g. P₁-Pₙ as mentioned in SC). It is these that give the object the kinds of causal resources to have the effects characteristic of a disposition’s typical manifestation(s). The same, then, can be said on such an account for a person, qua agent, and her abilities; the deeper explanation of her abilities, and most notably the ones constituting her free-will ability, will be by virtue of certain intrinsic properties possessed by the agent.

It is open to the new dispositionalists to develop their respective accounts in terms of responsiveness to reason, as both Vihvelin and Smith appear to do (though neither explicitly use the label “reasons-responsive theory”). For a relevant range of reasons, counterfactuals like SC could be constructed to help underwrite claims about the spectrum of responsiveness that Fischer and Ravizza attempted with their MRR—in particular, the spectrum of reasons to which an agent is both receptive and reactive. So, take Jones, and

---

20 Smith’s (2003) essay is written in a more general, programmatic style. He does not develop the kind of detailed proposal that either Fara or Vihvelin does, but it is clear from his exposition that he is open to the development of his view in ways are amenable to Fara’s or Vihvelin’s efforts.
for the moment set aside his acting in a Frankfurt example. Suppose for now
he is not in a Frankfurt example and that he shot Smith on his own and for
his own reasons. Suppose also that one of the reasons to which Jones would
have been both receptive and reactive in a way that would have resulted in his
not shooting Smith is if he were to have learned that Smith's child was with
him at the time. Were Jones to have learned of this, he would not have shot
Smith. But as it happens in the actual world, he learns of no such reason and
proceeds to shoot Smith. Here is a counterfactual, built by modeling it on
SC, that shows how it is that Jones is responsive to this reason, call it R1:

RR1: If Jones were to become aware of R1, and if Jones retained during
the relevant duration of time intrinsic agential properties P1-Pn, and if
Jones were not interfered with in a way that would impede the casual
efficacy of those properties, then Jones would not shoot Smith.

By compiling a collection of counterfactuals like RR1, say RR1 through
RRn, for a spectrum of reasons R1 through Rn, the new dispositionalists
could establish that an agent like Jones is MRR in just the fashion that
Fischer and Ravizza have been careful to work out.

Thus far, I have tried to show, in admittedly only broad brushstrokes,
how it is that by relying upon counterfactuals like RR1 through RRn, the
new dispositionalists could make use of their account of abilities to develop
a theory of reasons-responsiveness. As I have explained, it is open to them
to build the theory in such a way that it is very much like Fischer and
Ravizza's proposal for MRR but for the fact that their account would be
agent-based, not mechanism-based. But how is it that on their view the
truth of counterfactuals RR1 through RRn also underwrites the ability to do
otherwise in a way that shows how an agent retains this ability even in a
Frankfurt example? This is what is needed for them to put the nail in the
coffin of source compatibilists. Here their rationale is simple. Take Jones
whom we pulled from the context of a Frankfurt example, and take RR1
through RRn. Now place Jones back in a Frankfurt example. All of RR1
through RRn, the new dispositionalists will tell us, remain true. Like Fischer
and Ravizza, to assess the truth of RR1 through RRn they too must go to
worlds in which P1-Pn are allowed to operate unfettered, and this will also
require factoring out the presence of a counterfactual intervener like Black.
But according to them, this is perfectly consistent with Jones's possession of
the ability to do otherwise when he acts on his own. How so? When Jones
acts in a Frankfurt example, according to the new dispositionalists, Black
plays the role of a fink. When Black remains inactive, because he in no way
alters or impedes the efficacy of the pertinent agential properties such as
P1-Pn, Jones retains the ability to do otherwise—just like the vase sitting on
the shelf remains fragile despite that fact that, if toppled, a sorcerer would
turn it to stone. Of course, were Black to intervene, Jones would lose that ability. But Black remains passive (like the sorcerer). So, when Jones acts freely, according to the new dispositionalists, and contra Fischer and Ravizza, he retains the ability to do otherwise.

Is the new dispositionalists’ effort to undermine Frankfurt’s argument successful? No. As Clarke (2009: 339–42) is careful to point out, the problem with the new dispositionalists’ reply to source theorists is not that they fail to identify an ability that an agent like Jones retains in a Frankfurt example. Indeed, because when Jones acts on his own, none of his relevant dispositions are disturbed, there clearly is a kind of ability that he retains. But this ability can be described more carefully as a general capacity to do otherwise, or instead as a general ability to do otherwise that can, for instance, be exercised outside the context of Frankfurt examples, or when the agent is not asleep, or when she is not tied up, and so on.21 Nevertheless, as Clarke remarks:

> [T]here apparently are abilities that Jones lacks, because of Black’s readiness to intervene. Though Jones might have the capacity to act otherwise, the circumstances are not friendly to his exercising that capacity, and it may fairly be said that it is not up to him whether he exercises it, or that he does not have a choice about whether he does so. (2009: 340)

Furthermore, Clarke notes (341), Fischer, in articulating Frankfurt’s argument, is careful to point out that what is in dispute is whether under the circumstances an agent can chose and do otherwise (Fischer, 2002: 304). The salient point is that the new dispositionalists are only able to claim victory in refuting Frankfurt’s argument if the ability that they contend remains for an agent in a Frankfurt example is the same ability that is in dispute in debates about free will and moral responsibility—and it is this ability at which Frankfurt’s argument is aimed.

Precisely the same point applies regarding the question of whether an agent can be reasons-responsive while acting within the context of a Frankfurt example. If, as Fischer and Ravizza suppose, reasons-responsiveness requires that an agent be able to react otherwise in response to sufficient reasons to do otherwise, then what is in question is whether under the circumstances of a Frankfurt example, an agent is able to react as required. Recall (from Section 3) Fischer and Ravizza’s qualification regarding the counterfactuals in virtue of which MRR is confirmed; they do not involve worlds that are accessible to the agent (1998: 53). In the context of a Frankfurt example, it is not accessible to an agent like Jones to react otherwise in response to different reasons and, while in the presence of Black, not shoot Smith.

21 Ann Whittle makes a similar point in her insightful (2010).
I have argued that a mechanism-based account of reasons-responsiveness is not a viable option because it requires a principled basis for mechanism individuation. Regrettably, this leads to insuperable problems fitting mechanisms for the degree of plasticity called for by proposals like MRR. But if the reasons-responsive theorist is forced to return to an agent-based theory, is it possible for her to fit such a thesis into the constraints of Frankfurt examples, and thereby account for source freedom? According to Fischer and Ravizza it is not; Frankfurt examples rule out the ability to do otherwise, and agent-based reasons-responsive theories require that agents who are suitably reasons-responsive be able to react, and so do, otherwise. As I have explained, in light of this assumed conflict, the new dispositionalists are positioned to claim that the winner is reasons-responsiveness. According to them, what must go is a commitment to Frankfurt’s argument and the presumption that source freedom is freedom enough. But, the new dispositionalists can reason, this is all to the good, since their analysis of the free-will ability in terms of dispositions is, quite independently of any commitments to a reasons-responsive theory, sufficient to show that in a Frankfurt example an agent is able to do otherwise. The problem with this rejoinder is that the new dispositionalists have it wrong; they have not refuted Frankfurt’s argument. What is at stake in a Frankfurt example is whether—while in the context of a Frankfurt example—an agent is able to exercise an ability to do other than as she does. And nothing in the new dispositionalists’ playbook speaks to that issue.

So, assuming a commitment to Frankfurt’s argument and a focus on source freedom, is it possible to develop a reasons-responsive theory without relying upon the problematic notion of mechanisms? I think it is. The single proposition standing in the way of fitting an agent-based reasons-responsive theory to the contours of source freedom is that an agent in a Frankfurt example is not reasons-responsive because, being unable to do otherwise, she is thereby unable to react otherwise when given sufficient reason to do so. I think this proposition should be rejected. I now wish to propose a simple solution to this puzzle. An agent who acts freely in a Frankfurt example, I shall argue, is suitably reasons-responsive despite being unable to react otherwise when given sufficient reason so to do. I offer two points in support of my claim.

First, the fact that an agent in a Frankfurt example is not able to react otherwise in response to sufficient reasons does not exhaust the resources
available to show that, when she acts unimpeded, she, the agent, is reactive to sufficient reasons to do otherwise. Being reactive to sufficient reasons to do otherwise, I now propose, is not the same as being able to react otherwise in response to sufficient reason to do otherwise. In a Frankfurt example, an agent who is suitably reasons-responsive is reactive to sufficient reasons to do otherwise even if, due to the presence of a character like Black, she is unable to react otherwise.

To illustrate this point, consider again Jones outside the context of a Frankfurt example. Again suppose he shoots Smith for his own reasons. In this case, suppose that his actual reason for shooting Smith is revenge. Smith harmed Jones’s family, and now, as Jones sees it, it’s payback time. Let us also assign Jones a pattern of receptivity and reactivity that, we can stipulate, satisfies the conditions of MRR. Jones is receptive to reasons R1 through Rz, and is reactive to reasons R1 through Rn. Assume that the latter is a subset of the former in such a way that Jones’s degree of reactivity is weaker than his degree of receptivity. Recall that reason R1 as mentioned above is that Smith’s child is with him, and were Jones to learn of this, he’d be receptive to this as a sufficient reason not to shoot Smith, despite his actual reason of revenge, and he would react accordingly, not shooting Smith. And so it would be for the entire spectrum of reasons R1 through Rn, but not for the reasons ranging Rn+1 through Rz. For this latter range, Jones would be receptive to these reasons as sufficient for not shooting Smith, but he’d shoot Smith regardless. Imagine, for instance, that reason Rn+1 is that shooting Smith would cause Jones’s mother to be disappointed with her son. While Jones would recognize this as a sufficient reason for not shooting Smith, sorry son that he is, he’d not be reactive to it; he’d shoot Smith for his own reasons of revenge, despite the grief he’d knowingly cause his dear mum. Jones, given this set up, for the range of reasons R1 through Rn is both reactive to sufficient reasons to do otherwise, and would react otherwise in light of such reasons.

Now reinsert Black. Note that Black need not interfere with Jones were Jones to face up to any of the reasons within the spectrum of Rn+1 through Rz. Jones is not reactive to these reasons in the sense that they would not deter him from acting on his own reasons for shooting Smith. So Black is able to leave Jones to function on his own. Should Jones consider the reason Rn+1, that his mother would be disappointed with him if he shot Smith, he’d shoot Smith on his own anyway. Hence, Black can leave well enough alone. But as for the reasons R1 through Rn, were any of these reasons to become relevant to Jones’s practical context, Black would intervene. In being prepared to do so, clearly Black makes it the case that Jones is unable to react otherwise in response to these reasons. But it is nevertheless true that Jones is reactive to this spectrum of reasons R1 through Rn in just this
sense: for each reason, were it to become salient for Jones, it is not the case that, given his own intrinsic agential condition, he would act on his own reasons of revenge for shooting Jones. He is at least reactive to reasons in this manner. In reacting to this spectrum of reasons (at relevant possible worlds), he, by virtue of his own agency and his own reasons of revenge, is not the cause of shooting Smith. This is so despite that fact that he is unable to act otherwise and thereby avoid shooting Smith. While acting within the context of a Frankfurt example, Jones is reactive to this spectrum of sufficient reasons to do otherwise, R₁ through Rₙ, despite the fact that he is not able to react otherwise. His being reactive in this context consists in his not persisting in acting on his own reasons of revenge to shoot Smith in light of reasons such an R₁ (Smith’s child is with him).

Drawing upon the metaphysics of causation, Carolina Sartorio has made important contributions to our understanding of source compatibilist accounts of freedom, one of which is relevant to my current proposal. In threat-cancelation scenarios, Sartorio (forthcoming) points out, causation is not transitive along the path of a single causal chain. And the contrary to fact scenarios in Frankfurt examples—that is, the scenarios in which the intervener becomes activated—are set up as threat-cancelation scenarios. To explain: Jones, given his own reasons for shooting Smith, is such that, in the presence of certain reasons, such as R₁ (Smith’s child is with him), he would not shoot Smith. Suppose that, indeed, Jones is in the presence of this reason, R₁. This creates a threat to the event of his shooting Smith; it’s the sort of thing that’s liable to lead to Jones not shooting Smith. The presence of this threat, as a link in a causal chain, then causes the intervener to bring it about that Jones shoots Smith. Hence, there is a causal chain from Jones and his relation to R₁ through to his shooting Smith. But because of the way this causal chain unfolds, Jones, given the way he is and his own reasons prior to the intervention, does not cause his shooting of Smith. Why? Because, the way he is disposed to respond to reasons like R₁ creates a threat to his shooting Smith. So, despite the fact that there is a causal chain from Jones to the event of his shooting Smith via Black’s manipulation, Jones—just as he is as an agent, given his intrinsic properties—is not the cause of shooting Smith. Thus, as Sartorio makes clear, the difference between actual and counterfactual cases in Frankfurt examples is a difference in the causal role played by agents such as Jones. This, I maintain, helps cast light on the point I am at pains to emphasize here. The point has to do with the way Jones is reactive to the different kind of reasons that might bear on exercises of his agency. Some reasons, such as Rₙ₊₁, are such that, in their presence, Jones persists in being the right kind of agential cause of his action (of shooting Smith). Some, such as R₁, are
such that, in their presence, he does not persist in being the right kind of agential cause of his action.

The first point I am offering here bears some similarity to earlier failed efforts to defend PAP against Frankfurt’s argument by distinguishing acting on one’s own and not acting on one’s own, which seems to be a kind of alternative that cannot be expunged from Frankfurt examples (e.g. see Naylor, 1984). This strategy for defending PAP was handily rejected because the alternative of acting or not acting on one’s own in suitably structured Frankfurt examples were not alternatives that were themselves within the voluntary control of the agent (Fischer, 1994). Why doesn’t this apply to my current proposal? Here’s why: My focus on the distinction between an agent’s acting upon her own reasons in an actual scenario and her not acting upon her own reasons in alternative scenarios is not being used to underwrite any claim about an agent’s freedom to do otherwise. It is only being used to call attention to a fact about the mode of an agent’s acting in the actual-world scenario in which she acts on her own. The only reason that in the alternative scenario the counterfactual intervener causes Jones to act as he does is because, given Jones’s state, just as he is as an agent, he would not act on his own reasons of revenge for shooting Smith were these other reasons also salient. Granted, in the alternative scenario in which the intervener forces him to act, Jones does not do anything intentional to make it so that he does not act on his own reasons to shoot Smith; this scenario is not within the scope of his voluntary control. But there is no claim here that the degree of reactivity displayed by Jones in this range of counterfactuals affords Jones anything like a freedom to do otherwise. These counterfactual scenarios are not alleged to be accessible to Jones from the actual world. But, I say, they still display that, given the way he is in the actual world, he is in some manner reactive to pertinent reasons to do otherwise.

Perhaps some will find my argument here too thin. They might demand that being reactive to sufficient reasons to do otherwise requires that, in the presence of the relevant range of reasons, the agent’s own responsiveness-grounding resources have to be directly causally involved in the mode of differential reactivity. And when Black the intervener is in the driver’s seat—that is, when Black has taken over and is generating the action—that’s just not what is going on. In fairness, I grant that this is a reasonable source of resistance. However, in my own estimation, those inclined to press it in this context are fixed upon an alternative possibilities model of

---

22 Thanks to both Derk Pereboom and Brandon Warmke for raising this worry.
freedom, which the current proposal is meant to reject. Nevertheless, there is a distinct way of arguing that an agent in a Frankfurt example is suitably reactive to reasons, which I shall now explore.

Second, return to the complex counterfactuals which the new dispositionalists exploit for the purpose of advancing a compatibilist theory of the ability to do otherwise. Recall this one, where “R₁” names the reason that Smith’s child is with him:

RR₁: If Jones were to become aware of R₁, and if Jones retained during the relevant duration of time intrinsic agential properties P₁-Pₙ, and if Jones were not interfered with in a way that would impede the casual efficacy of those properties, then Jones would not shoot Smith.

As noted, to test for the truth of RR₁, we have license to go to worlds in which a counterfactual intervener like Black is missing. This is similar to the way that Fischer and Ravizza test counterfactuals regarding the dispositional properties of the mechanisms of action. The reason that factoring out Black in these contexts is felicitous is because it is what is needed to test the dispositional properties up for consideration. Now, the new dispositionalists might be wrong to think that counterfactuals such as RR₁ can be used in the service of giving an adequate account of the dialectically relevant ability to do otherwise. And Fischer and Ravizza might be wrong to think that we should draw on similar counterfactuals to account for the dispositional properties of mechanism rather than agents. But the counterfactuals themselves, such as RR₁, and in particular a sequence of them, say in a patterned form of RR₁ through RRₙ help to make what is in essence the same basic point I was at pains to emphasize early on (Section 2). These counterfactuals aid in demonstrating that when an agent like Jones acts as he does on his own,

23 Ginet (2006: 235–6) makes a similar point when advising Fischer and Ravizza to give up the mechanism component of their theory and instead just consider the reasons-responsiveness of agents.
the act he actually does perform is itself a reaction or a response to the actual conditions in which he acts. Furthermore, what these counterfactuals collectively help to underscore is not merely that Jones was indeed the cause of his so acting, but that the manner of his causing his act was sufficiently sensitive to reasons. 24 In this way, when an agent in a Frankfurt example acts, she is, so to speak, being suitably reactive in acting as she does, and so is exercising a kind of source freedom, a kind that can be characterized in terms of responsiveness to reasons.

One residual worry about the second point I offer here concerns cases in which, intuitively, an agent acts freely, but in which there are conditions intrinsic to her own agency such that these conditions play the role of a fink or a mask analogous to the role played by an extrinsic counterfactual intervener like Black. For instance, suppose that if Jones were not about to shoot Smith for the very particular sane reasons he had, then rather than Black intervening, some psychotic episode would unfold, leading to his shooting Smith anyway. Or we could instead run the case in terms of some latent phobia that would arise in the case of certain reasons but not others. If we hold fixed all of the intrinsic properties constituting Jones, we’d have to include these psychotic-constituting or phobia-constituting properties as well. In such cases, the relevant counterfactuals would come out false as applied to the agents on the proposal currently on offer. But if one were able to identify just certain features of the agent, as mechanisms, and hold these fixed, one might still be able to account for responsiveness. Clearly, these kinds of cases push back in the direction of a mechanism-based view. 25

24 An anonymous referee for OUP has raised the following insightful objection: Here I advocate an account of an agent’s reactivity by claiming that the pertinent counterfactuals help to reveal the manner of the agent’s causing her action. But, as I myself have noted above, Fischer and Ravizza themselves contend that they do not mean to reify the notion of a mechanism, and all they mean by appeal to an agent’s mechanism is the “manner that an action was caused.” Why, the referee asks, is my proposal any different from Fischer and Ravizza’s? This is an especially keen question, but the answer, to my mind, is telling in a way that speaks against Fischer and Ravizza and on behalf of my agent-based proposal. Even if Fischer and Ravizza mean to pick out no more than “the way an action is brought about,” they are at least committed to the existential claim of there being such a way, and to holding that fixed when testing it for responsiveness to different reasons. My proposal is that what gets held fixed for the pertinent testing is the agent. Now, either Fischer and Ravizza will grant that they mean the same thing, in which case their appeal to mechanisms really is the functional equivalent of the notion of agency itself and they might as well drop talk of mechanisms altogether, or they will insist that they mean to fix on something narrower than the agent. In that case we have just the very difference I have been at pains to bring out in this essay.

25 Stephen Kearns, Nicole Smith, David Sosa, and Jada Strabbing each pressed this challenging concern in slightly different ways.
Previously, it was precisely because of cases like these that I had been convinced that reasons-responsive theorists could not recover an agent-based theory while remaining committed to source compatibilism. But upon reflection, I do not think that these sorts of cases should carry the day. Admittedly, they are problematic. However, there are two quite different strategies for resisting their threat to the current agent-based proposal for explaining reasons-reactivity. One is to resist what seems intuitive in such cases. It’s not entirely clear that in such cases the agents really do act freely even when the pertinent psychosis or the phobia is inert. The agent’s “success” in acting uninfluenced by the troubling latent condition seems too fluky. The agent, after all, seems highly constrained in responding to just the reasons she does, so much so that her own agential resources impede her from responding to relevant patterns of reasons. A very different strategy would be to consider whether, when specifying the agent-constituting intrinsic properties that are to be held fixed in counterfactuals like R1, there is a principled way to rule out those constituting the psychosis or the phobia. The rough idea would be to treat these conditions as in some way alien or distant from those ingredients constituting the agent’s identity, or her real self.26 This is a promising avenue worth pursuing, though I’ll not do so here. It is enough, it seems to me, just to make clear that cases such as these pose a threat to the current proposal, but that there are avenues for addressing them which allow preservation of this basic claim. When a range of counterfactuals like RR1 through RRn are true of an agent, these do help to show how it is that when she acts as she does, she is reacting to reasons, even if she cannot react otherwise.

For a long while now, at least since the appearance of Fischer and Ravizza’s excellent and highly influential *Responsibility and Control* (1998), it has been assumed that the only way to marry a reasons-responsive theory of freedom to the lesson learned from Frankfurt’s argument is to forgo an agent-based theory. The way to account for source freedom in terms of reasons-responsiveness, I had previously thought, is by shifting to a mechanism-based theory. In light of the two preceding points, combined with the difficulties involved in accounting for mechanism individuation, I believe that this is mistaken. I propose an agent-based, reasons-responsive compatibilist theory of source freedom.

26 I am indebted to Michael Bratman for this point. He drew upon it to point out that this gives us reason to consider joining the resources of a reasons-responsive strategy with those of mesh theories like Frankfurt’s. The latter seek to make sense of something like the boundaries of the “real self” as narrower than the full spectrum on an agent’s psychological profile. (Though Bratman then raised the question of whether such a joint venture should be regarded as a friendly or instead a hostile takeover.)
APPENDIX I: WHY NOT A MESH THEORY RATHER THAN A REASONS-RESPONSIVE THEORY?

In the wider philosophical community, perhaps the best known source compatibilist account of freedom is *not* a reasons-responsive theory. It is Frankfurt’s own, which he expresses in terms of acting freely and of one’s own free will (1971). Frankfurt developed his position by attending to a fitting relation between higher-order and lower-order desires. An agent acts of her own free will when the first-order desire moving her all the way to action is the one that, at a higher order, she identifies with and so most wants to act upon. Here, the further details do not interest me. What does is the general strategy. Frankfurt’s is one version of a mesh theory because it accounts for freedom in terms of a harmonious mesh between different sub-systems within an agent’s overall mental economy. On Frankfurt’s approach, an agent’s freedom consists in the well-functioning relation between different orders of desire. When these different elements “line up,” the agent acts unencumbered, and so she acts freely. But when, for instance, what Sally most wants is for her desire not to drink the beer to win out, and when, despite that, her first-order desire to drink the beer leads her all the way to action, Sally acts from an unharmonious mesh. She is, in a sense, impeded or encumbered by her own psychological constitution and so, according to Frankfurt, is not free. The system issuing in her actions is out of alignment with the system that at a higher order constitutes what she most wants.

When reflecting upon cases in which an agent acts from a harmonious mesh, it is easy to see how that agent’s freedom, so construed, could easily be slipped into a Frankfurt example. So long as her action does actually issue from the mesh, there is no reason to be concerned about alternatives to what an agent does do. All that matters in such a case is that her well-functioning mesh operated unfettered. Thus, mesh theories such as Frankfurt’s fit seamlessly into the actual-sequence demands of Frankfurt examples. Why then explore the prospects of a reasons-responsive theory at all? Why not opt for some variation on a mesh theory?

In my estimation, all mesh theories, not just Frankfurt’s, face a deep problem for those committed to source compatibilism, one that I suspect is insurmountable. Mesh theories might comfortably capture conditions sufficient for acting freely: When an agent acts from a harmonious mesh, she acts freely. But in many cases in which an agent acts from an unharmonious mesh, mesh theories either generate the wrong results, or are instead unsatisfactorily silent—at least this is so for those committed to source compatibilism. Mesh theories generate the wrong results when they treat acting from a harmonious mesh as also necessary for

---

For other efforts to develop a mesh theory, see Dworkin (1970), and Watson (1975). Bratman (2007) has explored similar ideas. For a more detailed critical discussion of mesh theories, see McKenna (2011).
acting freely; they are unsatisfactorily silent when they refrain from accounting for an agent’s freedom in such cases.

To explain, consider the case of Sally above. Grant that on Frankfurt’s view, acting from a harmonious mesh is not only sufficient but also necessary for free action. Maybe Sally acts from a freedom-compromising compulsion in drinking that beer; suppose she is a full-blown, hopeless alcoholic. If so, there is no threat to Frankfurt’s mesh theory; it yields the result that Sally does not act freely insofar as she fails to satisfy the necessary condition at issue. But maybe instead Sally freely acts upon her lower-order desire as in opposition to the higher-order preferences with which she identifies. Suppose she also judges it best not to drink the beer. This is a classic case of weakness of will, understood in terms of freely acting contrary to one’s better judgment. A mesh theory such as Frankfurt’s generates the wrong result in a case like this. It has it that Sally does not act freely when in fact she does.

A viable option for the mesh theorist is to claim that the necessary condition which she proposes is merely an ability to act from a harmonious mesh. This would allow the mesh theorist to handle the distinction between exercises of free, weak-willed agency and failures that are due to freedom-undermining compulsion. The weak-willed agent, it can be argued, did not act from a harmonious mesh, but it remains true that she had the ability to do so; she could have done otherwise. Obviously, this reply will not work for any mesh theorist who, like Frankfurt, is also committed to source compatibilism, since it relies on claims about leeway freedom.

It is open to the mesh theorist who is committed to source compatibilism to retreat by weakening the reach of her theory. She could give up any proffered necessity condition and claim that she is only offering sufficient conditions for free action in those cases involving acting from a harmonious mesh. These conditions can be shown to be compatible with determinism without in any way relying upon any assumptions about leeway freedom. After all, this is all that is strictly required for her to answer the metaphysical challenge of the incompatibilist: the incompatibilist contends that determinism is incompatible with free action; she has produced cases of free action from a harmonious mesh that are compatible with determinism. Case closed. The objection to mesh theories currently under consideration concerns a refutation of the necessary conditions they propose. Why can’t the mesh theorist simply retreat by relinquishing this necessary condition?

Such a retreat strikes me as an unpromising dialectical tactic. A mesh theory that remained silent with respect to acting freely from an unharmonious mesh would leave unaccounted for a wide swath of (putatively) free actions. A credible theory of freedom should be able to account for these sorts of mundane cases of agency. And so a theoretical lacuna would invite further attempts to account for freedom in these cases. But then, once these kinds of cases are accounted for in ways that would be satisfying to compatibilists, it will only be natural to consider whether the proper account can also explain the freedom present in the cases involving a harmonious mesh. If so, then the work of accounting for free action will be
done more systematically with these alternative resources. Hence, the work
done by the mesh theory will be rendered otiose. This is, in fact, how I see the
explanatory resources of a reasons-responsive theory of freedom.28

APPENDIX II: ONE POINT WHERE FISCHER
AND RAVIZZA REQUIRE AN ACCOUNT OF
MECHANISMS

To illustrate a place where, without an independent rationale for doing so, Fischer
and Ravizza are too quick to allow sameness of mechanism to fall their way, consider
their contention that reactivity is all of a piece (1998: 73). They offer this proposal
in response to the incompatibilist worry that in the actual world, assuming a
deterministic context, when a blameworthy agent fails to act on the moral reason
to do otherwise, the agent cannot react to this reason. After all, holding fixed
the past and the laws, when that reason is present in the nearest possible world
(the actual one), the agent does not react to it. Fischer and Ravizza attempt to
ward off this objection by arguing that an agent’s reacting differently to a
different reason to do otherwise in another possible world is sufficient to
establish that she is able to react differently to any reason to do otherwise. Of
course, if this were so, she would be able to react differently to the moral reasons
that were present in the actual scenario, even though she did not. It is in this
context that Fischer and Ravizza (1998: 73–4) consider the following
incompatibilist-friendly challenge to their ‘reactivity is all of a piece’ thesis
(which I paraphrase here):

Is it not possible for the same mechanism to get more energy or focus from different incentives?
If it were, then the fact that an agent would react to some reasons to do otherwise, does not
mean that she is able to respond to all reasons to do otherwise, and so it does not mean that she
is able to react differently to the actual moral reason to do otherwise that was present to her at
the time of action.

If mechanisms were to behave this way—that is, in a way that disconfirms that
reactivity is all of a piece—then Fischer and Ravizza would be forced into a difficult
dialectical corner. They would be forced to take head on the incompatibilist worry
that, in a deterministic context, a blameworthy agent in the actual situation in which
she acts is not able to react to the moral reasons that were in fact present at the time.

I do, however, think that there is a way for mesh theories and reasons-responsive
theories to form a mutually beneficial alliance. The basic idea would be that a mesh
theory could be used to account for the nature of agency more generally (rather than free
agency). In particular, on the proposal I am considering, only agents able to adopt
higher-order attitudes towards their motivations can even be regarded as candidates for
having the kind of complexity for free agency as specified by something like MRR. Note
this might very well afford reasons-responsive theorists the resources to handle the sorts
of problems alluded to above. (See especially note 26).
They would have to show directly that, contrary to what the incompatibilist claims, such an agent is able to react differently to the particular moral reasons that were present, or they would have to commit to the claim that an agent could be blameworthy for failing to act on moral reasons that she was not able to act upon.29

To this challenge, they counter that in those cases in which an agent reacts differently to some incentives only because she acquires more energy or focus, it is natural to say that it is because a different mechanism of action is operative (74). This is one particular point where, I contend, Fischer and Ravizza are too quick to allow sameness of mechanism—or, rather, in this case, difference of mechanism—to fall their way (McKenna, 2001: 98–9). They claim that it is natural to think of the pertinent mechanisms as they propose. But it is hard to see how it is natural. It’s not clear in the first place what a mechanism of action is. It’s not a term of folk psychology that has accrued enough use to allow for much in the way of intuitive baggage. Why can’t a mechanism of action be reactive under, say, extreme pressures, but fail in normal contexts? After all, literal mechanisms, like thermostats, behave this way all the time, especially when they begin to fail. Indeed, what is the exception is to find an actual mechanism—a manmade gizmo of some sort—that does not respond differentially under extreme pressures, that does not have foibles and limits, and so on.

Fischer has responded to my objections by likening the problem I raise here to generality problems in other areas of philosophy. A certain level of generality, and thus lack of specification, is acceptable in other domains of inquiry, such as ethical theory. So too, Fischer argues, is a degree of generality acceptable when theorizing in terms of mechanisms of agency (Fischer, 2004: 169). Fischer also points out that it is unreasonable to demand of a successful theory that all of its elements are fully analyzed (168). Thus, he resists the burden of offering any “purely ‘principled’ account of mechanism individuation—an account that did not at some level appeal to intuition” (167–8). He thus remains committed to relying exclusively on appeal to intuitions in response to difference cases.

While I sympathize with Fischer’s thoughtful reply, it is not enough to put the objection to rest. Requesting some principled basis of mechanism individuation is not the same as demanding a full analysis of mechanisms—one that is “purely” principled. And a degree of generality is of course acceptable in theorizing in areas such as this one. But when the degree of generality and an exclusive reliance on intuition will not help adjudicate differences between cases crucial to assessing the theory, then demanding some principled basis for settling the dispute is only dialectically fair and reasonable. This is fully consistent with accepting Fischer’s contention that appeal to intuitions ought to enter the scene at some level.

29 I have defended the former of these two options (2005), while Pereboom (2006) has advised Fischer and Ravizza to opt for the latter. Watson has also commented on this issue (2001, as appearing in 2004: 300–1). He suggests that on pain of consistency, Fischer and Ravizza should opt for the latter option.
REFERENCES


