This is the penultimate version of an essay published in *Freedom and Determinism,* Ed. Joe Keim-Campbell, Michael O'Rourke, and David Shier (MIT Press) 2004.

Moderate Reasons-Responsiveness, Moral Responsibility, and Manipulation

Todd R. Long,
California Polytechnic State University

I may be a bird in a cage, but at least it's your cage.—Old 97's

Frankfurt-type examples have proven to be quite helpful in our thinking about the relation between freedom and moral responsibility. Harry Frankfurt designed his own examples to show that an agent can be morally responsible even if that agent could not have done otherwise.[1] Following Frankfurt's lead, many philosophers have developed their own Frankfurt-type examples to extend our understanding and sometimes to exercise our intuitions about the limits of moral responsibility and freedom.[2] In this essay, I join this enterprise by providing a novel Frankfurt-type example designed to reveal the limits of moral responsibility, given a new theory of responsibility developed by John Martin Fischer and Mark Ravizza. Although I am sympathetic to the mechanism-based theory of moral responsibility that Fischer and Ravizza (hereafter, "F&R") argue for in Responsibility and Control, I believe there is trouble for the way in which they apply their theory to Frankfurt-type examples. I will show that F&R's theory handles my test cases appropriately, but accepting my conclusion requires us (and them) to deny the seemingly plausible intuition that certain cases of rather severe external manipulation rule out moral responsibility.

In what follows I first summarize F&R's notion of guidance control with an eye toward explaining their preferred notion of reasons-responsiveness (namely, "moderate reasons-responsiveness"). Second, I show how F&R want moderate reasons-responsiveness to hook up with the relevant deliberative mechanism. Third, I offer a new Frankfurt-type example that both challenges F&R's own application of their theory to Frankfurt-type examples and threatens a certain appealing intuition that some people have about the relation between moral responsibility and external manipulation.

I. Guidance Control and Reasons-Responsiveness

In Responsibility and Control (hereafter, *RC*), F&R argue that, in the free will debate, *guidance control* is the freedom-relevant condition that is sufficient for moral responsibility.[3] Guidance control does not require the sort of *regulative control* that depends upon the existence of genuine alternative possibilities. Although regulative control often goes hand in hand with guidance control, F&R hold that moral responsibility requires only guidance control. Thus, we are morally responsible for our actions so long as we have guidance control, even if we lack regulative control. Pointing out that guidance control is the sort of control we find illustrated in

[1] See Frankfurt.
[2] For some of the works that have influenced my thinking, see Della Roca, Mele and Robb, Wideker, and Vihvelin.
[3] See Fischer and Ravizza (*RC*, chapters 1-2). See also Fischer (*The Metaphysics of Free Will*, chapters 7-8).

the Frankfurt-type actual-sequence examples, F&R conclude that guidance control is what ultimately grounds our ascriptions of moral responsibility. This is a familiar compatibilist claim. What sets F&R's theory apart is that F&R ascribe guidance control to an agent insofar as that agent's deliberative mechanism is appropriately responsive to reasons. Hence, their theory hinges on a satisfactory account of the deliberative *mechanism* operant in agents. I will return to mechanisms shortly.

Another significant feature of F&R's notion of guidance control is their distinction between moral responsibility, on the one hand, and praiseworthiness or blameworthiness, on the other. F&R characterize their notion of moral responsibility as a "Strawsonian view" that "allows for moral responsibility for "morally neutral" behavior. For instance, one can be morally responsible for simply raising one's hand" (*RC*, 8). All that is required for moral responsibility is that one be "an *appropriate candidate* for the reactive attitudes" (*RC*, 7). So, one can be morally responsible for an action without being an actual target of a reactive attitude. Nevertheless, moral responsibility is *connected* to praise and blame; after all, one would not even be a candidate for praise or blame unless one were morally responsible, in F&R's sense, for one's actions. We can summarize the key point here by saying that although guidance control (and, thus, moral responsibility) does not necessarily imply praiseworthiness or blameworthiness, the latter do imply the former.[4]

F&R refine their notion of guidance control by connecting it to the notion of *reasons-responsiveness*.[5] Their idea is that we are morally responsible for an action when the *mechanism* that issues in that action is reasons-responsive in an appropriate way. We can get a quick handle on the concept of reasons-responsiveness by thinking about contrasting cases. Consider Bill, who has a neurological disorder, which causes him to throw his head back violently every few seconds. Clearly, Bill lacks guidance control for his violent head movements. F&R claim that there is some physical mechanism issuing in Bill's behavior that has nothing to do with responding to reasons; therefore, we do not judge him to be morally responsible for his head movements. Now consider Amy, who is deliberating about whether to give money to a worthy charity; as she deliberates she may consider reasons for and against giving the money. Suppose she decides to give the money. Since her giving the money is not induced by a process such as a neurological disorder, a psychotic impulse, or brain manipulation, F&R hold that Amy's action is issued in by a mechanism that *is* responsive to reasons, and thus we think it proper to praise her for her charitable action; hence, she is morally responsible for her action.

---

[4] I thank Michael McKenna and Joseph Keim Campbell for bringing to my attention that F&R also mention an epistemic condition as necessary for moral responsiblity. Though providing no analysis of the condition, F&R say that it is meant to capture the "Aristotelian intuition" that "in order to be praiseworthy or blameworthy, a person must know (or be reasonably expected to know) what he is doing" (RC, 12). To use their example, if you were to run over a kitten hiding under your car but were ignorant of the kitten's location, then, intuitively, you would not be morally responsible for its death. After noting this epistemic condition, F&R don't mention it again but rather assume that it is met both in the examples that they discuss and in their claim that guidance control is sufficient for moral responsibility. In my CASEs 2-3 from section III below, I also assume that F&R's epistemic condition is met. It is possible that some epistemic condition would make trouble for my test cases, but I find nothing in F&R's discussion of their epistemic condition that implies that there is trouble for my cases; moreover, I doubt that there is a plausible, fully worked-out epistemic condition that would make trouble for my test cases.

[5] Here I summarize Fischer's discussion of reasons-responsiveness (*MFW*, 161-4). See also F&R (*RC*, 41-6).

F&R settle on a version of reasons-responsiveness that they call *moderate reasons-responsiveness* (or, **MRR**).[6] The following schema provides a fuller explanation of **MRR**:

MRR[7]: An agent's responsibility-relevant mechanism *K* is moderately reasons-responsive iff

(1) <u>*K* is regularly receptive to reasons, some of which are moral</u>; this requires (**a**) that holding fixed the operation of a *K*-type mechanism, the agent would recognize reasons in such a way as to give rise to an understandable pattern from the viewpoint of a third party who understands the agent's values and beliefs; and (**b**) that some of the reasons mentioned in (1.a) are moral reasons; and

(2) <u>*K* is at least weakly reactive to reasons</u>; this requires that the agent would react to at least one sufficient reason to do otherwise (in some possible scenario), although it does not follow that the *agent* could have responded differently to the actual reasons; and

(3) <u>*K* is the agent's own</u>; being the agent's own means "taking responsibility" for *K*; this requires that the agent (**a**) sees herself as the source of her behavior (which follows from the operation of *K*); and (**b**) believes that she is an apt candidate for the reactive attitudes as a result of how she exercises her agency in certain contexts; and (**c**) views herself as an agent with respect to (3.a) - (3.b) based on her evidence for these beliefs.

F&R hold that **MRR** is sufficient for moral responsibility.[8] Claims of sufficiency are customarily tested with specific cases. Any case in which an agent's mechanism is **MRR** but the agent is *not* morally responsible for an action issued in by that mechanism will count as a counterexample to F&R's claim of sufficiency. I will provide test cases in section III. But first a word about mechanisms.

II. <u>F&R's Human Deliberative Mechanisms</u>

Determining which of various mechanisms is actually operant in a given case is an important issue for F&R's account. What are the necessary and sufficient conditions for a mechanism to be the relevant mechanism in a given case of deliberation? F&R offer little help:

> We cannot specify in a general way how to determine which mechanism is "the" mechanism that is relevant to assessment of responsibility. It is simply a presupposition of this theory as presented here that for each act, there is an intuitively natural mechanism that is appropriately selected as the mechanism that issues in action, for the purposes of assessing guidance control and moral responsibility. (*RC*, 47)

---

[6] See *RC*, chapter 3. I will use **MRR** to stand for both "moderate reasons-responsiveness" and its cognate form, "moderately reasons-responsive".

[7] Although this schema does not appear in *RC*, I have done my best to reproduce faithfully the conditions of **MRR** as F&R present them. The schema contains quotations or near quotations of various claims made by F&R that are meant to explain what they summarize as follows: moral responsibility "requires that an agent act on a mechanism that is regularly receptive to reasons, some of which are moral reasons, and at least weakly reactive to reasons" (*RC*, 82), and also that the agent *takes responsibility* for the mechanism (*RC*, 242-3). Strictly speaking, on F&R's theory, condition (3) is not a condition of moderate reasons-responsiveness but is rather an additional necessary condition for guidance control. However, for ease of presentation I speak of **MRR** as encompassing all three conditions. This does not affect F&R's theory in any relevant way.

[8] F&R make this clear in numerous passages in *RC*. Here is a typical one: "Return to our account of moral responsibility for actions. On this account, an agent is morally responsible for an action insofar as he has guidance control of it, where guidance control consists in the action's issuing from the agent's own, moderately reasons-responsive mechanism" (*RC*, 230). As I mention in footnote 4, F&R make such claims on the assumption that their epistemic condition is met.

Nonetheless, F&R provide examples intended to help. One of their examples, which is about a case of direct brain manipulation, is useful for my present project:

> Consider, again, our claim that certain sorts of "direct manipulation of the brain" rule out moral responsibility. When the scientists intervene and manipulate the brain in a way which is not reasons-responsive, this is sufficient to undermine an agent's moral responsibility for action. Similarly, when forms of subliminal advertising, hypnosis, brainwashing, and so forth yield physical mechanisms that are not reasons-responsive, they rule out moral responsibility. Thus the theory that associates moral responsibility with actual-sequence reasons-responsiveness can help to explain our intuitive view that certain sorts of direct interventions in the brain are inconsistent with moral responsibility. (*RC*, 48-49)

Here is the crucial idea that F&R want us to get: any **MRR** mechanism that operates to issue in my actions is always a *different* mechanism from any one that would operate if my brain were monkeyed with to the point that I am no longer morally responsible for my actions. Hence, a brain-implanted mechanism that issues in an action would always constitute a different mechanism from an ordinary practical reasoning mechanism that issues in an action.

So far, so good. However, F&R also think that there is a close connection between the distinction we've just noted and Frankfurt-style examples. Specifically, they hold that the mechanism operant in an actual-sequence case is always to be distinguished from the mechanism operant in an alternative-sequence case. Consider Fischer's discussion of such cases in *The Metaphysics of Free Will* (hereafter, *MFW*):

> In a Frankfurt-type case the kind of mechanism that actually operates is reasons-responsive, although the kind of mechanism that would operate in the alternative scenario is not. In the Frankfurt-type case discussed above (in which Jones votes for Clinton on his own and Black does not actually intercede), Jones' action issues from the normal faculty of practical reasoning, which we can reasonably take to be reasons-responsive. But in the alternative scenario, a different kind of mechanism would have operated—one involving direct electronic stimulation of Jones' brain. And this mechanism is not reasons-responsive. Thus, the actual-sequence mechanism can be reasons-responsive, even though the agent is not reasons-responsive. (He couldn't have done otherwise.) . . . Clearly, . . . it is crucial to distinguish between the kind of mechanism that operates in the actual sequence and the kind of mechanism that operates in the alternative sequence (or sequences).[9] (*MFW*, 163)

I want to make two points about Fischer's discussion here. First, let's consider a couple features of this alternative-sequence case: (1) There is brain manipulation that leads to a particular outcome; and (2) The agent would have done otherwise had there been no manipulation. Now, one might think that what underlies our intuition that the agent lacks moral responsibility in such a case is some kind of manipulation principle that addresses those features. As we have seen, F&R hold that a brain-implanted mechanism issuing in an action is a different mechanism from

[9] I use Fischer's earlier account here because of the detail. Here is the parallel account by F&R in *RC*: ". . . in a Frankfurt-type case . . . the kind of mechanism that actually operates is reasons-responsive, even though the kind of mechanism that would operate—that is, that does operate in the alternative scenario—is not reasons-responsive. . . . But in the alternative scenario, a different kind of mechanism would have operated—one involving direct electronic stimulation of [the agent's] brain. . . . And this alternative sequence mechanism is not reasons-responsive" (38-9)

an ordinary practical reasoning mechanism. F&R say that this is just intuitively the case, but perhaps we can explain this intuition by drafting a manipulation principle, which addresses the two factors we just noted about alternative-sequence cases:

> **MAN**: If a person *S* performs an action *A* at time t_n only because another agent *B*, acting completely unbeknownst to *S*, has directly manipulated *S*'s brain earlier at t_1 then *S* is not morally responsible for performing *A*.[10]

We might think that F&R would happily endorse such a principle, for it helps to explain the intuition that a person is not responsible for an action if her brain has been directly manipulated in a certain way. More specifically, it helps to explain the appealing intuitive view that a person is not morally responsible for performing an action if the action is such that he would have refrained from performing it if his brain had not been directly manipulated by an external source. This is, after all, true of Fischer's Frankfurt-style case, wherein the agent is morally responsible in the actual-sequence but is not morally responsible in the alternative-sequence; and **MAN** would explain why Fischer concludes the quote above by saying that it is crucial to distinguish the mechanism operant in the actual-sequence case from the mechanism operant in the alternative-sequence case.

This leads me to the second point I want to make. Clearly, it is crucial to make this distinction only if it is true that the mechanism operant in the alternative-sequence case is always a different mechanism from the one operant in the actual-sequence case. But since F&R give us so little to go on concerning what a relevant mechanism is, or how to differentiate mechanisms, I see no reason to think that the mechanism that operates in the actual-sequence of a Frankfurt-type example *must* be different from the mechanism that operates in an alternative-sequence of a Frankfurt-type example. As I will show in section III, we can tell a Frankfurt-style alternative-sequence story in which the mechanism issuing in the action is the very *same* mechanism as the one operating in the actual-sequence case. The basic idea that I will develop there is that the results yielded by the human deliberative mechanism will vary depending on its inputs. It is very plausible to think that differing inputs (e.g. reasons) will yield differing outputs (e.g. actions). And in their latest account, F&R *grant* that their concept of the relevant deliberative mechanism in a given case involves the idea that the same mechanism may operate on a *variety* of inputs:

> . . . 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. . . .
>
> The picture here is of one kind of mechanism with different "inputs." (*RC*, 73-74)

F&R assume that in any Frankfurt-style example, the mechanism in an actual-sequence story is always a different one from the mechanism in an alternative-sequence story (call this F&R's distinction thesis). But, given that the same mechanism can yield different outputs depending on its inputs, we can conceive an alternative-sequence case as one in which the only difference between it and the actual-sequence case is the inputs—the mechanism remains the same.

[10] I thank Tom Senor for calling to my attention, though not necessarily endorsing, a version of **MAN**.

In *MFW*, Fischer says very little in support of his claim that the two mechanisms cannot be the same. Fischer suggests only that the mechanisms are different in the two cases because the *causes* are different; but that alone is not a good reason to accept his claim. After all, a light can be turned off by a human hand's flicking a light switch, and the same light can be turned off by falling plaster that strikes a light switch. Here we have two different causes for the light's going out, but we have one and the same mechanism (the light switch). In *RC*, F&R provide no additional reasons for thinking that the two mechanisms cannot be the same; rather, they just appeal to intuition to have us think that they must be different.[11] I contend that, given the dearth of explanation F&R provide about mechanisms, it can be just as intuitive to think that some actual-sequence and alternative-sequence scenarios involve the same, not different, mechanisms.

III. <u>The Test Cases</u>

In this section I will develop a basic Frankfurt-style actual-sequence example and then show three possible types of ways to get a different outcome in an alternative-sequence case. We will want to treat these examples as test cases for F&R's thesis that **MRR** is sufficient for moral responsibility, and we will want to find out what light these examples shed on F&R's distinction thesis. The basic story, which takes place in Germany during the 1930s, will involve some setup.[12] Here is the basic story:

Schmidt is a high-ranking official in the German government just prior to Hitler's rise to dictatorial power. Having been raised by parents with a deep inner moral sense and resolve, Schmidt learned from an early age the difference between right and wrong. As he developed, so did his moral wisdom. Suffice it to say that Schmidt developed into a man of impeccable moral judgment.

Due to Schmidt's high rank in the German government, he is a member of a secret group that will decide by vote whether or not Hitler is to be given supreme power. Schmidt has known Hitler for many years. The two became friends when each was a budding young politician. During those early years, they shared with each other their dreams and ideas about how to build a better Germany. Schmidt was impressed with both the ingenuity of Hitler's ideas and the deep moral sense that those ideas displayed. From the things that Hitler said to Schmidt during those early years, Schmidt had very good reasons to think that Hitler would be an honest, forthright, and morally upstanding asset to Germany's future. However, several years before the vote is to take place, Schmidt begins to hear disturbing stories about Hitler. From reliable sources, he hears that Hitler has intentions of building up a master race of pure-blooded Aryans, driving non-Aryans out of the country, and perhaps killing those who do not leave. Schmidt develops deep doubts about Hitler's moral character

Schmidt's role in the decision is to cast the deciding vote, should the voting members reach a deadlock. The members do reach a deadlock, so Schmidt is called upon to cast the deciding vote. Schmidt deliberates. Waiting in the wings is Block, a malevolent being with extraordinary causal and predictive powers. Should it become clear to him that Schmidt is going to vote

[11] F&R repeatedly say the following sort of thing: "the actual-sequence mechanism is, intuitively, the agent's *own*, whereas the alternative-scenario mechanism is *not*" (228).
[12] This story is fictional, but it is supposed to express the *character* of the historical Hitler.

against Hitler, Block will take effective steps to ensure that Schmidt votes *for* Hitler. Otherwise, Block will do nothing. So, whatever Schmidt's initial preferences and inclinations, Block will have his way.[13]

We can divide the basic possibilities here into two: Schmidt may deliberate and vote *for* Hitler. Although this would be a surprising outcome given Schmidt's moral character and his serious misgivings about Hitler, this would be a possible case under F&R's account of **MRR**; for F&R want their theory of moral responsibility to be compatible with cases of "weakness of will".[14] Note that in this case, which I will deem the *Block/Schmidt actual-sequence*, Schmidt is both morally responsible and blameworthy.

The other basic possibility is that Schmidt is such that, were Block not to intervene, he would vote against Hitler.[15] Now let us consider what course of action Block will take, if he believes that Schmidt will vote against Hitler, to ensure that Schmidt acts as Block wants. There may be a multitude of actions that Block could take; however, as far as I can tell, there are only three *types* of action that Block could take. I will apply each of these three possible types in the following three cases respectively, each of which will be an example of what I will call an *alternative-sequence*.

<u>BLOCK/SCHMIDT CASE 1</u>: A minute before Schmidt casts his vote, Block becomes convinced that Schmidt is going to vote *against* Hitler.[16] Block goes into action: rather than monkeying with Schmidt's normal deliberative voting mechanism, Block bypasses it. Block manipulates Schmidt in some way such that Schmidt's voting for Hitler is not related in any way to the reasons that Schmidt has for or against voting for Hitler. We might imagine him doing something directly to Schmidt's nervous system that physically causes him to vote for Hitler.

This is, of course, the kind of action that F&R think would occur should Block believe that he must act; for, in this alternative-sequence case, the mechanism that issues in Schmidt's voting *for*

---

[13] Readers may note that Schmidt and Block are counterparts to Frankfurt's Jones and Black, respectively. See Frankfurt, p. 162.

[14] Condition (2) of **MRR** allows for cases of "weakness of will". See *RC*, 42-44. In *MFW*, Fischer also wanted his theory to be compatible with cases of "actual irrationality" (166-8), but it is unclear whether F&R maintain this claim in *RC*.

[15] I put this as I do for two reasons (and I thank Michael McKenna for bringing this issue to my attention): First, I want to avoid the kind of criticism that David Widerker and others have brought against designing Frankfurt-type examples that both grant libertarian free action and rule out alternative possibilities. Such examples would have it that the agent at least *starts to choose* before the intervenor intervenes. The concern is that if some kind of non-determined act is allowed in the setup of a Frankfurt-type example, then there is no reason to think that the agent (Schmidt, in my cases) *cannot* do otherwise, regardless of the actions of the intervenor (Block, in my cases). So, as I characterize my cases, I don't assume libertarian free action on the part of Schmidt. This leads to my second point: F&R do not assume libertarian free action on the part of agents in Frankfurt-type examples either. Their view is that such examples show that alternative possibilities are not required for moral responsibility even if determinism is true. My goal is to follow their lead to see where it gets us. If it turns out that there is some general problem along these lines, then my cases might be problematic as well. However, I think I avoid problems in this neighborhood by setting up the cases as I do; for, all I need is for it to be the case that every possible world in which Schmidt does not do what Block wants is a world in which Block intervenes. I need not tell a story about how Block does this. If someone demanded such a story, then it seems to me that luck on Block's part could account for it.

[16] Although in all three of my alternative-sequence stories Block goes into action one minute before he thinks Schmidt will cast his vote, nothing significant turns on Block's timing.

Hitler (i.e., the one that Block uses) is a *different* mechanism than the one that would issue in Schmidt's voting *against* Hitler. In the actual-sequence, Schmidt's action is issued in by a deliberative voting mechanism that is **MRR**. But in this alternative-sequence, Schmidt's action is ushered in by a different mechanism, a mechanism that has nothing to do with reasons-responsiveness. As F&R say, we would hold Schmidt neither morally responsible nor blameworthy for his voting for Hitler in this case. Furthermore, note that F&R are committed to thinking that this is the *only* kind of action that could occur if Block should intervene; for, as we saw earlier, F&R's view is that we cannot hold Schmidt morally responsible in this case precisely because the mechanism that operates is not the same as the one at play in the actual-sequence (which Schmidt *has* taken responsibility for).[17] However, I see no reason to think that this type of action is the only possible way for Block to accomplish his work. Cases 2 and 3 demonstrate other types of action that could get the job done.

<u>BLOCK/SCHMIDT CASE 2</u>: A minute before Schmidt casts his vote, Block becomes convinced that Schmidt is going to vote against Hitler. Block goes into action: he *adds* new inputs into the *very same* mechanism that is operant in the actual-sequence. Suppose that these new inputs come in the form of *reasons* for voting for Hitler. Block directly feeds into Schmidt's deliberative voting mechanism enough reasons, or reasons powerful enough, to ensure that Schmidt will vote for Hitler. Because Block has extraordinary predictive powers, he knows which reasons will ensure that Schmidt's deliberative voting mechanism yields a vote for Hitler, and Block provides those reasons as inputs for Schmidt's deliberative voting mechanism.[18]

I have two things to discuss about this case. First, note that F&R do *not* consider this case as a possible way for Block to accomplish his desire. Their emphatic claim, that whenever there is intervention (in an alternative-sequence) the mechanism is a different one from the mechanism that would have been employed had there been no intervention, commits them to thinking that this case is impossible. However, I do not see that this case is any less possible than the one that they prefer. Second, this situation prompts the key question: Is Schmidt morally responsible for voting for Hitler? According to F&R's account of **MRR**, the answer is "yes", for Schmidt deliberates appropriately in this alternative-sequence case; that is, as he comes to make his decision, Schmidt's deliberative mechanism is regularly receptive to reasons, weakly reactive to reasons, and the mechanism is his own. In short, Schmidt meets all the conditions of **MRR**. Moreover, the deliberative voting mechanism that Schmidt employs is the same mechanism as the one operant in the actual-sequence—*that* mechanism can yield varying outcomes in response to different reasons. So, despite Block's intervention, Schmidt is morally responsible.

Now, one might worry about whether Schmidt is, in fact, morally responsible in CASE 2. We may assume that, had Block not intervened, Schmidt would have voted *against* Hitler. Even though Schmidt chooses via his own deliberative voting mechanism, something seems wrong about saying that he is morally responsible for voting for Hitler. After all, we might think, if Block hadn't monkeyed with the inputs (reasons) that Schmidt had on his own,

[17] F&R also consider a case in which scientists don't directly cause the implantee to perform an action but rather the scientists implant a mechanism that is itself **MRR**, which issues in an action. Although this example is closer to my CASEs 2 and 3, it is still a CASE 1-type in virtue of the fact that the mechanism itself is implanted.

[18] Of course, in order to avoid psychic shock or noticeable incongruence, Block may have to inject more than *thoughts* into Schmidt's deliberative voting mechanism. I just intend the story to be one in which Block adds in addition to reasons whatever is minimally necessary to ensure a vote for Hitler.

Schmidt would have voted against Hitler. Let's pause to consider why we might be inclined to think that Schmidt is not morally responsible in this case. I think it is because we are inclined to think that our principle **MAN** (above) is true: according to **MAN**, Schmidt is *not* morally responsible, since unbeknownst to Schmidt, Block manipulated Schmidt's brain a moment before Schmidt acts; and had Block not manipulated Schmidt's brain, Schmidt would have voted against Hitler. It is these facts that drive the intuitive sense that Schmidt is not morally responsible.[19] It seems that Block's action goes so far as to amount to the manipulation we had in CASE 1, even if the manipulation here occurs in a different way than it did in CASE 1. So CASE 2 appears to be a counterexample to F&R's claim that **MRR** is sufficient for moral responsibility.

<u>BLOCK/SCHMIDT CASE 3</u>: A minute before Schmidt casts his vote, Block becomes convinced that Schmidt is going to vote against Hitler. Block goes into action: he *removes* inputs from the *very same* mechanism in Schmidt that would have operated if Block had not intervened.[20] The inputs that Block removes from Schmidt's deliberative voting mechanism are Schmidt's own *reasons* for voting against Hitler. Block knows which reasons to remove in order to ensure that Schmidt's deliberative voting mechanism yields a vote for Hitler. Suppose that these reasons are in the form of Schmidt's negative memories of Hitler during the last few years (the ones that point to Hitler's diabolical intentions). The idea is that Block removes only the relevant memories—he does not remove whatever counts as the deliberative mechanism.[21] So, when Schmidt deliberates during that moment before he casts his vote, the only memories he has of Hitler are from the early days when Hitler seemed to be a thoroughly good chap (perhaps Block would also leave in Schmidt's brain morally neutral memories involving Schmidt's associating with Hitler during the past few years). Thus, the only reasons that Schmidt has to go on as he deliberates are reasons to vote *for* Hitler.

Is Schmidt morally responsible for voting for Hitler in this case? According to F&R's account of **MRR**, the answer is "yes", for in coming to make his decision, Schmidt's mechanism meets all the conditions of **MRR**. Moreover, the very same deliberative voting mechanism is operant in this case as it is in the actual-sequence. *That* mechanism can yield varying outcomes in response to different reasons. Schmidt exhibits guidance control in that the mechanism that actually issues in his action is **MRR**. So, F&R are committed to thinking that Schmidt is morally responsible, despite Block's intervention.

Nevertheless, it may seem that Schmidt is not morally responsible in this case. Although it is true that Schmidt's mechanism is **MRR**, it seems plausible to suppose that Block has deprived Schmidt of some of his own reasons in such a way that he is not morally responsible for voting for Hitler.[22] Moreover, note that if we are inclined to think that Schmidt is not morally

---

[19] I am claiming neither that F&R explicitly endorse **MAN** nor that they are committed to it; rather, the point of **MAN** is to provide something other than mere intuition to explain why it may seem to us that Schmidt is not morally responsible in this case.

[20] This possibility was suggested to me by Allen Plug in conversation.

[21] Block may, of course, have to remove more than just some of Schmidt's memories in order to avoid psychic weirdness in Schmidt, but the story is just that Schmidt removes whatever in addition to the relevant memories is minimally necessary in order to ensure a vote for Hitler.

[22] In order to motivate the worry that Schmidt is not morally responsible in this case, consider that it works as an amnesia case might. Consider Chuck, who has lost all memories of the last five years. Chuck has been married for

responsible in this case, it is because we think that **MAN** (or something very much like it) is true. On **MAN**, Schmidt is not morally responsible, since he was not aware of Block's manipulation a moment before Schmidt acted, and he wouldn't have acted as he did if Block hadn't intervened.

IV. The Options for Fischer and Ravizza

What resources do F&R have to deal with these last two test cases? The answer will depend on whether they concede them to be counterexamples. If they do, then I doubt that they can fix their theory without denying that **MRR** is sufficient for moral responsibility. Here's why: the only way around my counterexamples seems to involve construing the relevant deliberative mechanism broadly enough that it includes the inputs (i.e., the reasons) as part of the mechanism. On this construal, F&R would be entitled to conclude that Schmidt is not morally responsible in CASEs 2 & 3, because Schmidt has not taken responsibility for a mechanism that includes inputs that Block has either directly added or excised. This reply may seem promising, but it turns out to be deadly to their theory; for F&R want a *general* account that allows us to discern when an agent is morally responsible. Thus, what they need are mechanisms that are able to yield varying outputs (actions) depending on the inputs (reasons). But, on the present proposal, which has it that the mechanism *includes* the inputs, what we get in each case is an invariant mechanism. But if each mechanism is invariant, then we don't have what is needed for a general theory. To their credit, F&R maintain that the mechanisms they are talking about can, in fact, deliver varying outputs in response to various reasons. However, this claim opens them to the kinds of counterexamples that I raise. Here is a more general point: no matter how broadly one tries to construe the mechanism (i.e., no matter how many inputs are included as part of the relevant mechanism), one must allow for inputs that are *not* included in the mechanism. We can then develop a counterexample that will exploit those non-included inputs. Hence, if they accept CASEs 2 & 3 as counterexamples, then I cannot think of any way to save their theory without dropping the claim that **MRR** is sufficient for guidance control, and thus for moral responsibility. So, one option in dealing with my counterexamples is for F&R to affirm **MAN** (or something like it) and add conditions to **MRR** so that the theory gets the cases right.

However, F&R need not accept my cases as actual counterexamples. They have another solution open to them that will save their theory. Here is the alternative solution: my cases are not

---

five years, and his wedding anniversary was yesterday. Although Chuck knows the woman who is his wife (for he remembers having dated her), he did not know—indeed, could not have known—that yesterday was his wedding anniversary. So Chuck not only failed to buy his wife anything for their wedding anniversary, but he also failed to acknowledge their wedding anniversary at all. Assuming that Chuck had not been told earlier about his impending wedding anniversary, it seems clear that no one, not even Chuck's wife, ought to hold him responsible for not doing something special for his wife on their wedding anniversary (even if she is a bit disappointed). Although Chuck's relevant mechanism is **MRR**, we might think that he is not morally responsible for his lack of appropriate celebratory action, precisely because his amnesia has deprived him of the use of reasons that he once had. The idea here is that he is in no way responsible for the loss of those reasons, and so he is not morally responsible for failing to act appropriately. Compare CASE 3 to the reasoning in Chuck's case: Schmidt is in no way responsible for the loss of his reasons to vote against Hitler, and so he is not morally responsible for voting for Hitler. However, as I argue in the main text, Chuck may be morally responsible for whatever actions he does perform, so long as he acts in an appropriate reasons-responsive way; just note that if he murdered his wife, he would be morally responsible. So long as we are careful to distinguish moral responsibility from praiseworthiness and blameworthiness, we should have no trouble seeing that Chuck is morally responsible for his actions, even though he is not blameworthy for failing to do anything special for his wedding anniversary.

counterexamples at all; Schmidt *is* morally responsible in CASEs 2 & 3, but he is not blameworthy. Although it is unfortunate that Schmidt's action put Hitler in a position to carry out his horrible, unspeakable acts, and although we may assume for the sake of the thought experiments that had Schmidt voted against Hitler, Hitler would not have carried out his reign of terror, nonetheless in each of the two cases Schmidt did the best he could given the reasons he had. Assuming that he had no special reason to doubt his own faculties, it is clear that in each case Schmidt had very good reasons to vote for Hitler. Moreover, Schmidt voted in consonance with his reasons. Given his reasons, it would have been irrational for him to vote against Hitler. These facts show that Schmidt is morally responsible for his actions.

So where do these considerations leave F&R? Either they can accept my examples as counterexamples, in which case they will need to go back to the drawing board in order to hash out the missing necessary condition(s) of **MRR**, or they can deny that my examples are counterexamples. If the latter, they will need to acknowledge a false assumption they seem to be under about Frankfurt style-cases. To make this point perspicuous, I want to draw out what I take my discussion to have shown about Frankfurt-style cases. The fundamental thing that Frankfurt designed these cases to show is that one can be morally responsible even if one couldn't have done otherwise. But the alternative-sequence cases are diverse. Using my characters as examples, the leading idea of an alternative-sequence case is that every possible world where Schmidt on his own does not do the thing that Block wants is a world where Block intervenes such that Schmidt does what Block wants. I have shown that we can divide the alternative-sequence cases into two subcategories, one of which nobody seems to have noticed. There is the kind of case in which Block intervenes in such a way that Schmidt is not morally responsible for his action—that is, the case in which Block does something to Schmidt that physically (or otherwise) directly causes Schmidt to vote for Hitler. This is the kind of case that F&R and others have countenanced. The other kind of case is one in which Block intervenes to bring it about that Schmidt does what Block wants, but the intervention is such that Schmidt does not lack moral responsibility for his action. My examples show not only that Block can manipulate in a way that eliminates moral responsibility, but they also show that Block can manipulate in a way that does not eliminate moral responsibility. Now, I suppose that everyone will agree that Block could have successfully manipulated Schmidt by more ordinary means—by providing him with efficacious newspapers, flyers, and other documents, or by arranging a large-enough bribe, or by some other means—and Schmidt would be morally responsible for his actions in each of those cases. The key point I have made is that the manipulative intervention that does not eliminate moral responsibility can be more severe than we might have thought. We are now in a position to state what I said above is a false assumption by F&R. They seem to assume that in all alternative-sequence cases, the intervenor's manipulation eliminates moral responsibility on the part of the agent. My examples show that this assumption is false.

There is another lesson to be drawn from my examples, which involve manipulation that does not rule out moral responsibility. This kind of manipulation can be so severe that it clashes with an intuition that many people have. That intuition is the driving force behind **MAN**, which says that if an agent is unknowingly manipulated in such a way that she performs an action that she wouldn't have performed had there been no manipulation, then she is not morally responsible for her action. Despite the fact that **MAN** may seem intuitively right, further reflection reveals that there is little to recommend it. Consider BLOCK/SCHMIDT CASEs 2 & 3. In each case, not

only did Schmidt meet the conditions of **MRR**, but he also intuitively did the morally responsible thing, given the evidence he had to go on. And if we knew the relevant facts about each case, we might even praise him for his action, even as we deeply regret or deplore the trick Block played on him and the consequences that followed. This shows that **MAN** is false, for Schmidt *did* perform an action at time t_n only because another agent, acting completely unbeknownst to Schmidt, directly manipulated Schmidt's brain earlier at t_1; and yet Schmidt *is* morally responsible for voting for Hitler, since he acted in consonance with his reasons in an appropriate way. The moral is that we are morally responsible for our actions when they are consonant with our reasons in an appropriate way (perhaps, when **MRR** is satisfied), but whether we are praiseworthy or blameworthy for our actions depends upon additional facts about the cases. Counterexamples to **MAN** are plentiful. Although the intuition behind **MAN** may be appealing, it does not survive scrutiny.

After doing this ground clearing, we are in a position to see that F&R's theory *does* get the cases right. It implies that when the deliberative process as a *whole* is directly manipulated the agent is not morally responsible for actions issued in by that process. This is exactly right. The problem is in F&R's assumption about how their theory applies to Frankfurt-style examples; for it does not, nor need not, imply that all alternative-sequence cases eliminate moral responsibility on the part of the agent.

V. Trouble with Intuitions

However, Fischer himself has responded to my examples by saying that intuitively Schmidt is not morally responsible in CASEs 2 & 3. Fischer infers from this intuition that the manipulation of the inputs makes the mechanism in those cases a different mechanism from the one operant in the actual-sequence case.[23] I find this response disappointing, for Fischer concedes that he has no principled way to distinguish the mechanisms in question. If **MRR** really is sufficient for guidance control, and thus for moral responsibility, as F&R claim that it is, then an objection to my examples stands a chance only if we have reason to doubt that Schmidt meets the third condition of **MRR**. However, I think it is clear that Schmidt can easily meet that condition. In my examples I suppose that Schmidt (a) sees himself as the source of his behavior, which follows from the operation of the relevant mechanism, that Schmidt (b) believes that he is an apt candidate for the reactive attitudes (in the manner F&R specify), and that Schmidt (c) views himself as an agent based on his evidence for (a) and (b). Nothing about Block's manipulation prevents Schmidt from meeting those conditions. To think that Schmidt fails to satisfy those conditions is to suppose that Schmidt would have to be responsible both for his prior evidence (or beliefs) and for the way in which he acquired that evidence (or those beliefs). But, surely this is to add too much to the notion of taking responsibility. Evidence (and reasons) supporting beliefs come to us from various directions and in many different ways. Some of it we are quite aware of, and some of it we are less aware of. Earlier I mentioned that nobody would balk at the suggestion that Schmidt could have been convinced to vote for Hitler by more ordinary external means (news reports, personal testimony, etc.). What is the relevant difference between manipulating the inputs in those ways and manipulating the inputs directly?

[23] Fischer made these claims during the question-and-answer session of my presentation of a version of this paper at the Inland Northwest Philosophy Conference on Freedom and Determinism, University of Idaho, April 2001.

Consider a case in which a group of conniving folk manipulate the inputs of someone's deliberative mechanism by affecting the information that person gets. Suppose a group of Young Republicans who work at a news press manipulate the inputs that Karen's deliberative voting mechanism receives by changing the text of Karen's newspaper every time something good is said in print about a Democratic candidate. Would this prevent Karen from being morally responsible for her voting? If you are inclined to answer "yes", then you owe us an explanation of why *any* of us are morally responsible for our acts of voting. After all, very few of us control the political reports or the political opinions that are propounded in our newspapers and other media; thus, most of us do not control the inputs our deliberative voting mechanisms receive. You can bet your bottom dollar that much of the political talk in newspapers is written with the intent of manipulating us to vote the way the authors want us to vote. On what basis should we judge ourselves as morally responsible for our voting but deny moral responsibility of Karen? My point is even more convincing when we turn our attention to advertising. The very livelihoods of advertisers depend on their ability to manipulate us into buying whatever marketers are selling. Advertising is a sneaky, subtle business, but are we to suppose that most of us are not morally responsible for our purchases? Despite the cunning powers of advertisers, surely most of us are not helpless victims of manipulation. Likewise for Schmidt in CASEs 2 & 3; he was no helpless victim. No one caused him to act by bypassing his deliberative voting mechanism. Schmidt actually weighed reasons in a regularly reasons-responsive way, and his deliberative voting mechanism was weakly reactive to reasons; moreover, that mechanism, which we may assume Schmidt used many times before, was one that he had clearly taken responsibility for. So, according to **MRR**, Schmidt is morally responsible in CASEs 2 & 3.

F&R want their theory to capture our intuitions about moral responsibility. I grant that judging Schmidt to be morally responsible strains the prereflective intuition that some people might have about the cases. However, it is fairly common for us to find that our prereflective intuitions about a matter are wrong, after reflection on the details; and, I have found that, after reflection, many people (philosophers and non-philosophers) judge the cases as I do. Even so, I am not inclined to hang my case on any claim about how many people have a particular intuition. Instead, I put forward the following evidence for your consideration: (1) F&R want variant mechanisms; that is, they want any **MRR** mechanism to be able to operate on a variety of inputs and to produce a variety of outputs; and (2) F&R explicitly distinguish between moral responsibility, on the one hand, and praiseworthiness and blameworthiness, on the other. Consideration of (1) reveals that it will not do to build all the inputs into an **MRR** mechanism, but the only way to make sense of Fischer's response to my examples is to think that Schmidt would have to be responsible for his input acquisition as well as for the mechanism itself. This amounts to building the inputs into the mechanism. Consideration of (2) reveals that F&R have the resources for their theory to get the cases right. Schmidt *is* morally responsible for voting for Hitler, although he is not blameworthy. If you are still unconvinced, suppose that Schmidt walked into the voting room and murdered the officials who voted against Hitler. I think that we would all judge him to be morally responsible for the murders, even if we knew every detail about Block's manipulation. But, then, what principled reason is there to deny moral responsibility of Schmidt for his voting?

In summary, given that F&R explicitly distinguish between any **MRR** mechanism and its inputs, given that they explicitly distinguish between moral responsibility and praiseworthiness and

blameworthiness, and given that they have no principled reason to judge Schmidt not to be morally responsible in CASEs 2 & 3, I think it is clearly more reasonable to judge Schmidt as morally responsible in those cases; furthermore, I think that this judgment is correct.

VI. Conclusion

F&R are aware that, in failing to provide effective differentiation parameters for mechanisms, they open their theory to philosophical worries:

> Clearly, on the approach to moral responsibility we have begun to sketch, it is crucial to distinguish between the kind of mechanism that operates in the actual sequence and the kind of mechanism that operates in the alternative sequence (or sequences). We must confess that we do not have any general way of specifying when two kinds of mechanisms are the same. This is a potential problem for our approach; it will have to be considered carefully by the reader. (*RC* 40)

After carefully considering the point, I conclude that BLOCK/SCHMIDT CASEs 2 & 3 reveal an *actual* problem for F&R's characterization of their theory with respect to Frankfurt-style examples. As I have shown, there may be no distinction between the mechanism that operates in the actual sequence and the mechanism that operates in the alternative sequence (or sequences). The problem, however, is not insurmountable, although my examples show that a reasonable solution requires F&R to give something up: either they must give up the claim that **MRR** is sufficient for moral responsibility, or they must give up their claim that **MRR** explains why all Frankfurt-style alternative-sequence cases yield agents who are not morally responsible for their actions. As I have shown, alternative-sequence cases need not yield agents who lack moral responsibility for their actions. **MRR** *does* rule out moral responsibility when manipulation is pervasive across the whole deliberative process, but it does not rule out moral responsibility when manipulation touches only the *inputs* to the deliberative process (even though the manipulation may be so severe that it conflicts with **MAN**). Since I think there is something right about connecting moral responsibility to reasons-responsiveness, I think it best to take the second horn of the disjunction. Not only is this solution plausible, but it also keeps F&R's theory of moral responsibility intact.[24]

Todd R. Long,
California Polytechnic State University

[24] Special thanks to Rich Feldman for many fruitful discussions about this paper. Thanks also to John Martin Fischer, Tom Senor, Michael McKenna, Joseph Keim-Campbell, Allen Plug, Greg Wheeler, Tony Roark, Ken Himma, and an anonymous referee for this volume, as well as participants in the Rochester Philosophy Society (2000), the 52nd Northwest Conference on Philosophy (2000), the American Catholic Philosophical Association—Western New York (2000), the Society of Christian Philosophers—Eastern Region (2001), the Philosophy Writing Seminar at the University of Rochester (2001), and the Inland Northwest Philosophy Conference on Freedom and Determinism (2001), for helpful comments on this paper.

Works Cited

Della Rocca, Michael, “Frankfurt, Fischer and Flickers”, *Nous*, Vol. 32, No. 1 (1998) 99-105.

Fischer, John Martin and Ravizza, Mark, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press, 1998.

Fischer, John Martin, *The Metaphysics of Free Will: An Essay on Control*, Oxford: Blackwell, 1995.

Frankfurt, Harry, “Alternate Possibilities and Moral Responsibility”, *Free Will*, ed. Derk Pereboom, Indianapolis: Hackett, 1997, pp. 156-166.

Ginet, Carl, "Comments on John Fischer, The Metaphysics of Free Will", *Journal of Social Philosophy*, Vol. 29, No. 2 (1998) 126-134.

Mele, Alfred R. and Robb, David, “Rescuing Frankfurt-Style Cases”, *The Philosophical Review*, Vol. 107, No. 1 (1998) 97-112.

Widerker, David, "Frankfurt's Attack on Alternative Possibilities", *Philosophical Perspectives*, Vol. 14, 2000, A Supplement to *Nous*, ed. James E. Tomberlin, pp. 181-201.

Widerker, David, "Libertarian Freedom and the Avoidability of Decisions", *Faith and Philosophy*, Vol. 12, No. 1 (1995) 113-118.

Widerker, David, “Libertarianism and Frankfurt’s Attack on the Principle of Alternative Possibilities”, *The Philosophical Review*, Vol. 104, No. 2 (1995) 247-261.

Vihvelin, Kadri, “John Martin Fischer, The Metaphysics of Free Will”, *Nous*, Vol. 32, No. 3 (1998) 406-440.