SOFTENING FISCHER’S HARD COMPATIBILISM

C. P. Ragland

Abstract: According to “hard” compatibilists, we can be responsible for our actions not only when they are determined by mindless natural causes, but also when some agent other than ourselves intentionally determines us to act as we do. “Soft” compatibilists consider freedom compatible with merely natural determinism, but not with intentional determinism (e.g., theological determinism). Because he believes there is no relevant difference (NRD) between a naturally determined agent and a relevantly similar intentionally determined agent, John Martin Fischer is a hard compatibilist. However, he argues for “historical” compatibilism by appealing to the intuition that certain manipulated agents are not responsible. By considering a new type of manipulation case, I show that Fischer’s appeal to ordinary intuitions about manipulation conflicts with NRD, so that he must choose between the two. The closing section explains why I think going “soft” is Fischer’s better option.

Compatibilism about free will and determinism comes in two varieties: hard and soft.1 To see the difference between these, we must distinguish between ‘intentional’ and ‘natural’ determination. An event is naturally determined just in case it must occur given past history and the laws of nature. An event is intentionally determined just in case it is causally necessitated by some agent’s deliberate choice to bring it about. Hard compatibilists, such as Harry Frankfurt (1986, 113–23) and Gary Watson, say that moral responsibility is compatible with both natural and intentional determination of the will. On their view, we can be responsible for our actions and choices not only when they are determined by laws of nature, but also when some agent other than ourselves is deliberately causing us to choose and act as we do; we can be “at once a full fledged free agent” and “a creature whose every move, every hope and scheme is a part of another’s plan” (Watson 2000, 69). Soft compatibilists, such as A. J. Ayer (1982) and Daniel Dennett (1984, 57–73), claim that freedom and moral responsibility are compatible with natural determination, but not with intentional determination. In their view, being completely under the control of another person is not compatible with being free and responsible. However, past history and the laws of nature are not agents capable of controlling us, and so in

their view natural determinism is no threat to moral responsibility.

This paper aims to bring out a tension in John Martin Fischer's writings about moral responsibility. While he is officially a hard compatibilist, his response to certain kinds of manipulation cases seems better suited to soft compatibilism. In section one, I will show why Fischer's "official position" is hard compatibilism. In section two, I summarize Fischer's account of the control condition for moral responsibility and present a manipulation counterexample to it. In sections three and four, I consider two ways that Fischer might try to respond to the counterexample, both of which try to preserve his hard compatibilist position. I argue that these two responses are not very plausible. This leaves Fischer with a dilemma: he must either make a counterintuitive claim about the manipulation case I set forth, or he must give up his hard compatibilism. I close by very briefly indicating what the latter course might look like, and why it should be acceptable to Fischer.

1 NEUTRALITY AND HARD COMPATIBILISM

Fischer's compatibilism rests in part on his conviction that our practices of holding one another morally responsible—of giving and receiving praise and blame, of feeling resentment, approbation, and so forth—do not depend for their justification on the truth or falsehood of certain physical or metaphysical theories. Regarding physical theories, Fischer says the following:

One of my motivating engines in trying to argue that determinism is compatible with responsibility is that I think that our view of ourselves as persons and as morally responsible agents should not hang by a thread. It should not be held hostage to the possibility that some consortium of scientists will discover and announce tomorrow that causal determinism is true. If they announced that, I would not stop holding people responsible. (MacDonald et al. 2000, 413)

Fischer believes we would be justified in holding some people responsible for their actions even if we knew for sure that physical determinism is true. From this he concludes that indeterminism is not a necessary condition for moral responsibility. Similarly, Fischer would not give up on responsibility if scientists proved that nature is not deterministic. Therefore, determinism must not be a necessary condition for moral responsibility, either. Fischer believes that moral responsibility is neutral with respect to physical determinism, in the sense that the truth or falsehood of such determinism is simply not relevant to whether people are morally responsible. Let's call this belief Fischer's "physical neutrality" intuition.

Fischer also has what I will call a "theological neutrality" intuition. He thinks that our practices of holding one another morally responsible should not "hang by a thread" on the answers to certain questions about the existence and nature

---

2 Fischer makes similar claims in Fischer et al. (2007, 44–45, 71).

3 Fischer calls his position "super-compatibilism" MacDonald et al. (2000, 413).
of God. For all we know, traditional theism may be correct: it could be “that God set up the world and knew in advance its entire history,” but such divine foreknowledge or preordination (if it exists) “does not eliminate our responsibility” (MacDonald et al. 2000, 415). This would be true even if God foreknew what would happen by determining it to happen. Just as moral responsibility is neutral with respect to physical determinism or indeterminism, so it is neutral with respect to the presence or absence of theological determinism.

Fischer thinks that if deterministic laws of nature do not threaten freedom, then it should not make any difference whether those laws are “free-standing” (as in a naturalistic metaphysical view) or were put in place to by an omniscient, omnipotent God to fulfill a providential plan (as in traditional theism). Fischer says:

I am beginning with the claim that causal determination in itself does not rule out moral responsibility . . . . Then I am suggesting that, if a process analogous to mere causal determination but initiated by God issues in human action, then the etiology should not in itself rule out moral responsibility either. (1994, 181)

These remarks commit Fischer to hard compatibilism. He holds that moral responsibility is compatible with divine causal determination, which is an instance of intentional determination. So Fischer must think that moral responsibility is compatible with at least some kinds of intentional determination of human choice—that is, he must be a hard compatibilist.

Notice the reasoning that leads Fischer to hard compatibilism. In the passage above, he compares two different scenarios: one in which a person’s action results from a merely natural deterministic causal process, and another in which it results from a causal process that is exactly similar to the first, save for being deliberately initiated by another person (God). Let us call these the “natural scenario” and the “intentional scenario” respectively. In the above passage, Fischer suggests that there is no relevant difference (for moral responsibility) between these two scenarios: if the causal sequence produces a morally responsible agent (and action) in one scenario, then it will produce a morally responsible agent in the other. Therefore, if the person is responsible for her action in the natural scenario, she will also be responsible in the intentional scenario. The crucial assumption here can be expressed as follows:

No Relevant Difference (NRD): given a given natural deterministic scenario (N) and an intentional deterministic scenario (I) involving impersonal mechanisms exactly similar to those in (N), an agent is responsible in (N) if and only if she is responsible in (I).

NRD says, in effect, that merely adding design to a naturally deterministic scenario, without changing anything else about it, should not change our assessment of whether agents are morally responsible in that scenario.

4 For another position similar to Fischer’s on this score, see Mele (1995, 186–191)
In holding NRD, Fischer is in good company. Many of the most important contributors to current free will discussions endorse it. Here are a few representative quotations:

It is hard to see what differences there could be between the natural and purposeful forms of determination that would be relevant to freedom and control. (Watson 2000, 67)

The causes to which we are subject may . . . change us radically, without thereby bringing it about that we are not morally responsible agents. It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberative manipulative designs of other human agents. (Frankfurt 2002b, 28)

Determinism is in no relevant manner any different than is this sort of [global covert non-constraining (CNC)] manipulation—it is just a different way to bring about the very same result. (McKenna 2004, 169)

The problem is to locate the relevant difference between the two that makes . . . CNC control . . . objectionable and . . . mere determination . . . not. This problem turns out to be more difficult than first appearances suggest . . . . Why should the distinction between CNC control and mere determination by natural causes make a difference in our powers, when the results are also the same? (Kane 1998, 68)

NRD is especially important to incompatibilists, because it is at the heart of manipulation arguments for incompatibilism. Such arguments are similar to Fischer’s reasoning above, but run in the opposite direction. They begin with the intuitive conviction that the person in the intentionally determined scenario is being manipulated, and therefore is not responsible for her actions. But there is no relevant difference between this manipulation scenario and a merely deterministic scenario. Therefore, mere determinism must be incompatible with moral responsibility.6

2 MANIPULATION AND MORAL RESPONSIBILITY

Though he is officially a hard compatibilist, Fischer is not satisfied with certain versions of hard compatibilism, particularly those offered by Gary Watson and Harry Frankfurt. Watson and Frankfurt offer “current-time-slice” theories of responsibility, according to which an agent has control over her action in virtue

5 McKenna here expresses a premise in a manipulation argument for incompatibilism, so he is reconstructing the views of others. However, later in the same article (pp. 178–9) he clearly endorses this premise himself.

6 Though he is not himself an incompatibilist, Mele’s “Zygote Argument” is a clear example of this reasoning (2006, 189). See also McKenna’s presentation of the “Manipulation Argument” for incompatibilism McKenna (2004, 169–70).
of responsibility-conferring features she possesses at the time of action (such as proper fit between her first and second-order volitions (Frankfurt 1971), or between “valuational” and “motivational” preferences (Watson 1982)) regardless of how she came to acquire those features. Fischer believes that any such theory can be shown inadequate by manipulation counterexamples involving covert non-constraining control (CNC).

A CNC-controller can somehow produce an agent’s entire motivational structure, including higher-order, evaluative desires. Victims of CNC “act in accordance with their own wants, desires, or intentions. Yet they are controlled nevertheless by others who have manipulated their circumstances so that they want, desire, or intend only what the controllers have planned” (Kane 1998, 65). A common example of a CNC-controller is a neurosurgeon who secretly places a “remote control” device in someone’s brain. The device allows him to physically alter the victim’s brain states, thereby causally determining her to perform certain actions for certain reasons.

In cases of CNC, it seems intuitively that the agent is not morally responsible for her action. Often, in response to CNC cases, paradigmatic hard compatibilists like Frankfurt and Watson simply “bite the bullet”: they deny our ordinary intuitions about such cases, and insist that if manipulated agents satisfy the right conditions, they can be responsible for their actions. As Watson puts it: “For the compatibilist, the constitutive conditions of free agency do not conceptually depend on their origins. In this sense, free and responsible agency is not an historical notion” (Watson 2000, 66).

Fischer is not satisfied by Frankfurt and Watson’s response to many CNC manipulation cases. For example, he says:

I think that manipulation cases are compatibilism’s dirty little secret.Compatibilists don’t like to admit that this is a problem. It is to Bob Kane’s and other incompatibilists’ credit that they have pushed us to confront cases of covert non-constraining control . . . . We compatibilists have to deal with this. In my view, honestly, Harry Frankfurt really has not addressed this problem. He has discussed it in different ways and in different places and it doesn’t add up to much. (Fischer 2000b, 390)

As this quotation suggests, Fischer has the intuition that in many manipulation cases, the manipulated agents are not morally responsible. Apparently reasoning as follows, he takes the manipulation cases as counterexamples to Frankfurt’s (or Watson’s) view:

(1) Intuitively, the manipulated agent is not responsible (despite meeting the sufficient conditions for freedom on the current-time-slice account).

(2) If the current-time-slice account were correct, then the manipu-
lated agent would be responsible.

So, (3) The current-time-slice account is incorrect.

Fischer concludes that moral responsibility must require some further conditions beyond those specified by a current-time-slice theory. In short, Fischer appeals to our ordinary intuitions about such CNC cases to argue that moral responsibility is a historical phenomenon: being responsible is not just a matter of having certain properties at a time, but of acquiring those properties in the right sort of way.\(^9\) As Michael McKenna has noted, Fischer’s argument here moves “from intuitions about these sorts of [manipulation] cases to historical conclusions” about the nature of moral responsibility (McKenna 2004, 185)\(^10\) and that is why Fischer and Ravizza call it an “intuitive idea that what matters, for moral responsibility, is not just the arrangement of mental ingredients, but how that arrangement is produced” (1998, 187; my italics).

In Responsibility and Control, Fischer and Ravizza maintain that moral responsibility requires “guidance control.” An agent exercises guidance control over an action just in case she satisfies both the following conditions:

Reasons-Responsiveness: the action is produced by a moderately reasons-responsive mechanism.

Ownership: the aforesaid mechanism is the agent’s own, in the sense that the agent takes responsibility for that mechanism.

Though much could be said about the first condition, our present purposes require focusing on the second, ownership condition.

As Fischer and Ravizza note, “the operation of a moderately reasons-responsive mechanism just prior to the action can itself be the product of a process that intuitively rules out moral responsibility.” Such a mechanism “could be ‘implanted’ by a scientist, or produced via direct electronic stimulation of the brain” (1998, 230). In other words, Fischer is willing to grant that a reasons-responsive mechanism could be produced and activated by covert non-constraining control. In such a case, he claims, the manipulated agent is not responsible because she fails to satisfy the ownership condition.

Briefly, an agent “owns” the mechanism that produces her action by coming to have the right sorts of beliefs about herself and her actions. A person takes responsibility for a mechanism just in case she meets the following three conditions: (1) she sees herself as the cause of the actions (and their consequences in the world) produced by that mechanism, (2) she believes that she is an apt candidate for other people’s reactive attitudes of resentment, gratitude, etc., in response to actions produced by the mechanism, and (3) her beliefs (1) and (2) are based in an appropriate way on the evidence available to her (1998, 197–239).\(^11\)

---

\(^9\) Just as, for example, genuine sunburn requires not only having irritated skin, but having it as a result of excessive exposure to sunlight.

\(^10\) See also McKenna (2004, 174–76); as McKenna notes, Haji and Mele offer similar arguments for historical compatibilism.

\(^11\) See also Fischer (2000b, 389–90), and MacDonald et al. (2000, 409).
According to Fischer and Ravizza, ordinary adults have normally taken responsibility for the mechanisms of practical reasoning and unreflective habit. They have not taken responsibility for the type of mechanism exploited by the controller in CNC cases (e.g., direct brain stimulation). Therefore, in CNC cases involving reasons-responsive mechanisms, the victim of control is not morally responsible because the mechanism producing the action is not her own (1998, 232–34).12

This ownership condition also helps explain our intuitions about what I call “fluke” cases. These are cases parallel to the CNC cases just discussed, but where the reasons-responsive mechanism originates not with a behind-the-scenes manipulator, but with some kind of strange or abnormal impersonal mechanism, like a brain disease or passage through “the Bermuda Triangle.” 13 Many have the intuition that agents whose behavior is causally determined by such sources are not morally responsible for what they do, even if the mechanism is reasons-responsive. Fischer and Ravizza can explain this intuition: a fluke mechanism is not the mechanism normally operative in practical deliberation or unreflective habit, so the agent will not have taken responsibility for it.

We can also imagine cases in which the CNC or fluke mechanism produces (in a reasons-responsive way) not only an action, but also the relevant cluster of beliefs concerned with taking responsibility. As Fischer and Ravizza put it, “it is conceivable that the individual’s view of himself as an agent and an apt candidate for the reactive attitudes be electronically implanted” (1998, 235–6). Fischer and Ravizza argue that in such cases, the ownership condition is not satisfied because the agent’s beliefs about himself are not properly based on evidence. The evidence condition on taking responsibility “is intended (in part) to imply that an individual who has been electronically induced to have the relevant view of himself (and thus satisfy the first two conditions on taking responsibility) has not formed his view of himself in the appropriate way” (1998, 236). This evidence condition requires further scrutiny.

Fischer and Ravizza stress that the ownership requirement makes their theory a “subjectivist” theory, in the sense that an agent “must have a certain view of himself in order to be morally responsible” (1998, 221).14 Given this emphasis on the agent’s own perspective, it is natural to understand Fischer and Ravizza’s evidence requirement along “internalist” lines. So understood, the evidence (on which the agent’s view of himself must be based) would be the agent’s immediate, subjective experience. On an internalist view, the justification of the relevant beliefs would not depend on factors in principle inaccessible to the agent, such as whether or not (from a third-person perspective) the beliefs are formed by a reliable mechanism. However, a CNC-controller or fluke cause could in principle implant not only the relevant beliefs, 15

---

12 For an interesting criticism of Fischer and Ravizza’s claims on this point, see Long (2004).
13 Mele discusses such cases (1995, 168–69).
14 In this respect, their view aligns with that of Strawson (1986).
15 More generally, as long as the conditions for taking responsibility involve only properties internal
but also the internal evidence on which those beliefs need to be based.\textsuperscript{15} Therefore, it seems that Fischer and Ravizza would want to avoid this internalist construal of the evidence requirement.\textsuperscript{16}

Furthermore, the following passage suggests that Fischer and Ravizza probably have an “externalist” evidence requirement in mind. They say:

For example, the child’s view of himself as an agent needs to be based (in an appropriate way) on his experience with the effects of his choice and actions \textit{on the world}. And his view of himself as an apt target for the reactive attitudes in certain contexts needs to be based on what his \textit{parents} have taught him and his broader experiences with the \textit{social practices} of (say) praise and blame. (1998, 213; my italics)

It appears that the child’s subjective experiences count as evidence (for purposes of taking responsibility) only if they are caused in the right way by the “external world.” To have evidence, the child needs to be interacting with a real physical environment and with real people. Therefore “memories” or “experiences” implanted by a CNC-controller or fluke mechanism—though indistinguishable from real evidence from the agent’s point of view—would not count as \textit{genuine} evidence.\textsuperscript{17} Therefore, although a CNC controller could manipulate a victim into \textit{believing} that she was responsible for her manipulated actions, she would not be really responsible, because her relevant beliefs about her own agency would not be properly based on evidence.

I think that Fischer and Ravizza’s evidence requirement, understood in an externalist fashion, successfully blocks the potential manipulation counterexamples discussed so far. However, we can imagine a further sort of manipulation case, and it is not entirely clear what Fischer would want to say about it. Consider, then, the case of “Manipulated Jimmy”:

Assume determinism is true. Jimmy has grown up under the careful supervision of Alex, a super-psychologist who knows enough about Jimmy to flawlessly predict how his environment will effect—by means of Jimmy’s own properly functioning faculties—his choices to the agent, and not the agent’s relations to the external world, it will be plagued by manipulation counterexamples of this type. For as Eleonore Stump (2002) has noted, if Cartesian dualism is false, then an “internal” process of taking responsibility—a process that results in one’s ownership of a mechanism—must have some kind of physical basis in a person’s brain. Therefore, by somehow producing this physical basis, a CNC-controller could make a person take responsibility for the mechanism of CNC-control.

\textsuperscript{16} My point here echoes Haji (2000, 397).

\textsuperscript{17} Long (2004, 157–61) presents some interesting counterexamples to Fischer and Ravizza’s theory, in which a manipulator somehow puts reasons into, or removes memories from, the evidential inputs that drive his victim’s ordinary mechanism of practical deliberation. I believe that a slight expansion of Fischer and Ravizza’s evidence requirement, understood in an externalist way, may give them the resources to block Long’s counter-examples. Suppose that the requirement for proper external grounding applies not just to the evidence that leads us to take responsibility for a certain mechanism, but more broadly to all evidence relevant to our practical deliberations (this is plausibly part of the standard epistemic condition for moral responsibility). In that case, Long’s manipulator could be seen as severing the proper evidential connection between his victim’s normal deliberative mechanism and the world, so that the agent does not count as responsible.
and behavior (Alex knows, for example, that if Jimmy is placed in environment type A, he will rebel against Alex’s control over his circumstances, but if he is in environment type B, he will continue to accept Alex’s governance; consequently, he maintains Jimmy in an environment of type B). As he grows, Jimmy takes responsibility in the ordinary way for his own reasons-responsive mechanisms of practical deliberation and unreflective habit. But over the years, Alex uses his environmental control over these mechanisms to groom Jimmy for performance of a specific terrorist act. When the time is ripe, Jimmy performs the act from his own mechanism of practical deliberation, and in such a way that he satisfies all the non-control-related conditions for moral responsibility.18

When we consider what Fischer would, or should, say about this type of case, it is easy to see the tension between his hard compatibilism on the one hand, and his use of manipulation cases on the other. This is especially true if we compare Manipulated Jimmy to the similar case of “Natural Jimmy”:

Assume determinism is true. By fate or chance, but not as the result of any person’s design, Natural Jimmy grows up in an environment that is physically exactly similar to the environment Manipulated Jimmy grew up in. Because of this environmental history and the laws of nature, Natural Jimmy performs the terrorist act.

As far as I can see, Fischer has two basic options for responding to the Jimmy cases. First, reasoning as he did in the theological case discussed above, Fischer could claim that because Natural Jimmy is morally responsible, Manipulated Jimmy must be as well. I will call this the “hard line” response. Second, he can say that Manipulated Jimmy is not responsible, and therefore presents a genuine counterexample to Fischer’s own account of moral responsibility. Reasoning as he did with regard to current-time-slice theories, Fischer could take Manipulated Jimmy’s intuitive lack of responsibility to show that his theory needs amendment.

3 THE HARD-LINE RESPONSE

In explaining his theological neutrality intuition above, Fischer started by assuming the correctness of his theory of moral responsibility. On that theory, an agent can be morally responsible when determined by merely natural causes. Fischer then reasoned that because there is no relevant difference between that naturally determined agent and an exactly similar theologically determined counterpart, the counterpart would also be morally responsible.

There is only one difference between the naturally determined agent and the

18The case of Manipulated Jimmy is similar (in different respects) to Cases One and Three of Derk Pereboom’s “Four Case” generalization argument for incompatibilism (2001, 110–22). It also resembles the case of Diana and Ernie which Alfred Mele uses as the basis for his “zygote argument” for incompatibilism (2006, 188–90).
theologically determined agent: the latter is part of another person’s deliberate
design, carrying out another’s plan. But this is also the only difference between
Natural Jimmy and Manipulated Jimmy. Therefore, it would not be surprising
for Fischer to respond to the Jimmy cases as follows: “Because he satisfies all
the conditions in my theory of moral responsibility, Natural Jimmy is responsi-
ble. There is no relevant difference between Natural Jimmy and Manipulated
Jimmy. So, Manipulated Jimmy is also responsible.”

This “hard line” response will strike many people as highly counterintuitive.
When we learn that Jimmy has been so thoroughly manipulated, can we blame
him for his crime? If we can, don’t we blame him hesitantly, with mixed feel-
ings? The unease we feel about blaming Jimmy is a significant problem for this
line of response. However, Fischer may have a way to soften the counterintu-
tiveness of his position, much as he did in response to the first case of Derk
Pereboom’s “Four Case” Argument.

Pereboom imagines Professor Plum, who was “created by neuroscientists,
who can manipulate him directly through the use of radio-like technology . . .
directly producing his every [mental] state from moment to moment” (2001,
112–3). He is as much as possible like an ordinary human, and the neuroscien-
tists manipulate him in such a way that his ordinary mechanisms for delibera-
tion are moderately reasons-responsive. One day, the neuroscientists cause
him—by means of his ordinary deliberative mechanism—to kill Mrs. White. Is
he responsible for the killing?

Somewhat surprisingly, Fischer answers “yes.” He notes that “Plum has
taken responsibility for the manipulation-mechanism” because it is “the mecha-
nism on which he always acts, and when an individual develops into a morally
responsible agent, he takes responsibility for his actual-sequence mechanisms,
even if he does not know their details” (2006, 232). Since Plum owns the mech-
nanism, and it is moderately reasons-responsive, Fischer judges that “he has a
certain minimal measure of control” (2006, 233) and so is morally responsible.

Given the radical nature of the manipulation to which Plum is subject, this
response is counter-intuitive—indeed, much more so than is holding
Manipulated Jimmy responsible. If Plum were responsible, it would be appro-
priate to blame him for his bad action. But how can we blame Plum when he is,
ineffect, a puppet of his manipulators, thinking and moving each moment only
at their whim? Aware of this problem, Fischer tries to block the counterintuitive
consequence of his position by challenging the connection between responsibil-
ity and blameworthiness. He says: “Professor Plum . . . is not blameworthy,
even though he is morally responsible” (2006, 233).

Fischer points out that being morally responsible for performing an action is
not by itself sufficient for deserving blame (or praise) for that action. Agents
who perform morally neutral actions are neither praiseworthy nor blameworthy,
even if they are morally responsible for those neutral actions. To be blameworthy,
one must not only be morally responsible, but also must have performed a
morally bad action. But even satisfying both these conditions may not be
enough to make one blameworthy, Fischer thinks. He says:
In my view, further conditions need to be added to mere guidance control [i.e. moral responsibility] to get to blameworthiness; these conditions may have to do with the circumstances under which one’s values, beliefs, desires, and dispositions were created and are sustained, one’s physical and economic status, and so forth. (2006, 233)

Fischer explains that even though Professor Plum is responsible and performed a bad action, he is not blameworthy because “the circumstances of the creation of his values, character, desires, and so forth” (2006, 233) do not satisfy these “further conditions” for blameworthiness. Clearly, what Fischer says about Plum, he could also say about Manipulated Jimmy, thereby blocking the counterintuitive consequence of taking a hard line on that case, as well.

I think that Fischer’s position here—that there are additional conditions for blameworthiness beyond moral responsibility for a bad action—runs counter to the standard way of defining “morally responsible” in the philosophical literature. Witness the explanation of that term in *The Stanford Encyclopedia of Philosophy*: “... to be morally responsible for something, say an action, is to be worthy of a particular kind of reaction—praise, blame, or something akin to these—for having performed it” (Eshleman 2009). Now, Fischer’s point about the moral status of an action is surely correct, so a more precise version of this definition would say something like “to be morally responsible for performing a morally bad action just is to be worthy of blame for having performed that action.” But the point remains that as “morally responsible” is normally used, the conditions for moral responsibility (whatever they may be) *just are* the sum total of those conditions that a person must satisfy to count as blameworthy for performing a bad action (or praiseworthy for performing a good one). For this reason, I think that if we intuitively consider Professor Plum *not at all* worthy of blame, then we think he has failed to satisfy some condition on moral responsibility. The same goes for Manipulated Jimmy.

Nevertheless, I think that Fischer may still have a way to square his “hard line” response with our intuitions. For while moral responsibility is (in my view) the only factor that determines whether agents are blameworthy for doing something bad, how blameworthy they are—the extent of the blame we can appropriately lay on them—may well be determined by additional factors of the sort Fischer indicates in the passage above. Indeed, this may be what Fischer is really driving at in the passages cited above, for he sometimes alludes to degrees of blameworthiness in his discussion of this issue. For example, he says that in Pereboom’s case, Professor Plum is morally responsible for killing Ms. White, even if he is “not blameworthy (or not fully blameworthy) for doing so” (2006, 233), and later he says that Plum is “most likely not blameworthy (or significantly blameworthy)” (2006, 235).

It is plausible that blame comes in degrees, and that it can be mitigated by the circumstances of an action or agent. For example, many people seem to think that slaveholders of old, though certainly worthy of very severe blame (let me not be misunderstood!), are not quite *as* blameworthy as those who capture and
enslave others today. The old-time slaveholders deserve less blame, perhaps, because they were psychologically formed in societies that sanctioned and encouraged slavery (whereas today’s slaveholders are without excuse). For various reasons, I do not want to wholeheartedly endorse this thought here, but it does have an initial degree of plausibility. And it opens the way for Fischer to say that the “further conditions” he specified above are conditions not on blameworthiness simpliciter, but on full blameworthiness. He could say that although Plum and Manipulated Jimmy are blameworthy, the conditions under which they were formed mitigate the blame appropriate to them: they are not very blameworthy, perhaps just barely blameworthy. Their very low level of blameworthiness explains why we feel queasy about blaming them to any significant extent. In this way, Fischer’s “hard line” response can be made to seem fairly intuitive.

Unfortunately for Fischer, if this strategy is successful (a matter on which I reserve judgment), its very success presents him with a further problem. Recall the dialectic of the foregoing discussion. Fischer’s theory entails that Manipulated Jimmy (or Professor Plum) is morally responsible. When it is objected that intuitively, Jimmy (or Plum) does not seem blameworthy, we have imagined Fischer responding roughly as follows: “Your intuitions only seem to conflict with my theory; for in stating this objection, you are forgetting that sometimes circumstances extraneous to moral responsibility can mitigate blame. Though Jimmy (or Plum) is responsible, he is only barely blameworthy, given the bizarre circumstances of his formation and ongoing manipulation, so there is no wonder that you hesitate to place significant blame on Jimmy (or Plum).” Now let us suppose that Fischer’s appeal to mitigating circumstances (or further conditions for full blameworthiness) succeeds in squaring his theory with intuition. The question arises: why couldn’t a “current-time-slice” theorist like Frankfurt or Watson use this same strategy—with equal success—in response to Fischer’s objections that their theories yield counterintuitive results? They too could point to “additional factors” about cases which would mitigate blame, thereby explaining away the apparent conflict between their theories and ordinary intuitions. In fact, Michael McKenna has already worked out this sort of defense of non-historical compatibilism. He articulates the general strategy as follows:

The nonhistorical compatibilist . . . can attempt to draw further differences in our judgments about [normal versus manipulated agents] beyond those regarding questions about their freedom and responsibility . . . . Focusing upon these further considerations might help lessen the counterintuitive appearance of the claim that [manipulated] agents . . . are just as free and responsible [as normal agents]. (McKenna 2004, 184)

It seems that Fischer faces a dilemma, then, concerning his “hard line” response to the Manipulated Jimmy case. On the one hand, he can grant that his theory yields counterintuitive results. This is a significant strike against the the-
ory, especially for a philosopher like Fischer, who aims to achieve a kind of “reflective equilibrium” between his theory and our pre-reflective moral intuitions (1998, 10–11). On the other hand, he can explain away the counterintuitiveness of his theory’s results by pointing to additional factors that mitigate blameworthiness. But if he does this, he takes much of the force out of his own arguments against non-historical theories of moral responsibility. Given this dilemma, it seems Fischer would do well to carefully consider the alternative way of looking at Manipulated Jimmy—as someone not morally responsible.

4 GOING SOFT

Notice that in the Manipulated Jimmy case, Alex controls Jimmy by means of his environment (both physical and social). This is the same type of control to which Robert Kane points in discussing B. F. Skinner’s Walden Two, a scenario in which “all the residents are conditioned from childhood by behavioral engineers to want and choose only what the conditioners have planned” (Kane 2000, 401). Kane argues that such conditioning vitiates responsibility for the residents of Walden Two, and uses this intuition as a starting point to build an argument for incompatibilism: since the residents of Walden two are not responsible, and there is no relevant difference between them and merely naturally determined agents, naturally determined agents also would lack responsibility for their actions.

In responding to Kane, Fischer seems to admit that the residents of Walden Two are not responsible (Fischer 2000a, 405). He quarrels with other assumptions of Kane’s argument instead. And as McKenna notes, not only Fischer, but other historical compatibilists like Mele and Haji generally “take it that global manipulation cases like . . . Walden Two provide powerful intuitive support for the historical component” of their theories (McKenna 2004, 174). If Fischer thinks that the residents of Walden Two are not responsible, then it would be natural for him to agree that Jimmy, too, is not responsible. This would be consistent with his tendency to try to accommodate ordinary intuitions about manipulation cases as far as possible.

But taking this “soft line”—granting that Manipulated Jimmy is not morally responsible—seems to present Fischer with two serious problems. First, it would seem to commit him to both premises of Kane’s manipulation argument for incompatibilism (or at least an analogue of Kane’s argument). For like Kane, Fischer accepts NRD: he thinks that if Manipulated Jimmy is not responsible, then Natural Jimmy won’t be either. Second, Jimmy seems to meet all the conditions for responsibility set out in Fischer’s theory. Therefore, if Jimmy is not responsible, then his case proves the following remark by Harry Frankfurt:

An agent who is being manipulated in ways that undermine moral responsibility can, according to the criteria that Fischer and Ravizza provide, act on a mechanism that is both suitably reasons-responsive and the agent’s own. Thus . . . their criteria do not satisfactorily identify the conditions upon which moral responsibility depends. (2002a, 61)
Saying that Jimmy is not responsible would seem to force Fischer to amend his theory.

I think that Fischer could try to avoid these problems (without amending his theory) by appealing to a distinction introduced by Harry Frankfurt. Frankfurt considers two different ways in which a CNC-controller, which he dubs the devil/neurologist (D/n), might determine his victim’s actions: either by constant, direct intervention, or by implanting a stable program into the victim. Frankfurt agrees that the first way would eliminate the victim’s freedom, but claims that the second way would not.

Frankfurt describes the first sort of intervention as follows:

The D/n manipulates his subject on a continuous basis, like a marionette, so that each of the subject’s mental and physical states is the outcome of specific intervention on the part of the D/n. In that case the subject is not a person at all. (1986, 119)

In this “marionette manipulation”, the victim’s actions and experiences do not genuinely cause one another because they are all directly caused by the D/n. The victim’s mental states relate to the D/n in much the same way that the states of creatures relate to God in Malebranche’s occasionalism: they can provide the true cause (God or the D/n) with a reason to act, but they themselves do not genuinely cause their apparent effects. Just as Leibniz insisted that merely occasional causes cannot be genuine substances (1989, 159–60), so Frankfurt claims that the D/n’s victim is rendered insubstantial: there is not enough to him, metaphysically, to count as a person.

Frankfurt spells out what we might call “program manipulation” as follows:

The D/n provides his subject with a stable character or program, which he does not thereafter alter too frequently or at all, and . . . the subsequent mental and physical responses of the subject to his external and internal environments are determined by this program rather than by further intervention on the part of the D/n. (1986, 119)

In such a case, Frankfurt maintains, the subject counts as a genuine person, and can be morally responsible for his actions (1986, 121).

Fischer clearly agrees with Frankfurt that there is a significant difference between marionette manipulation and program manipulation. He maintains that in cases where the victim is subject to direct brain manipulation from infancy, “that baby never becomes a person,” but that in cases where “superhuman intelligences . . . just start the world in such a way that they know it will play itself out as they wish it to,” with no intervention, the subjects could be morally responsible (MacDonald et al. 2000, 415). Following Frankfurt’s line in this way may allow Fischer to respond to both problems above, showing (1) why on his theory as it stands Manipulated Jimmy is not responsible, and (2) why his commitment to NRD does not force him into incompatibilism.

On the first point, Fischer could say that Jimmy does not really satisfy all of his theory’s conditions for moral responsibility. One such condition is that
Jimmy be a person. But in Responsibility and Control Fischer and Ravizza suggest that someone who is subject to continual manipulation (as is Jimmy or Professor Plum) cannot become a real person:

If an individual has been subject to continuous significant manipulation (of which he has been unaware) since he was quite young, then it appears that our account would imply that he has in fact taken responsibility for this sort of manipulation. But we would respond that such an “individual” cannot ever have developed into a coherent self. That is, under the envisaged circumstances, there is no self or genuine individual at all—from the beginning, there has been no opportunity for a genuine self to emerge and develop. (1998, 234–35, note 28)\(^{19}\)

Although Jimmy (unlike Professor Plum) acts on his own stable character, Fischer could perhaps argue that Jimmy is subject to a close analogue of marionette manipulation, and is therefore not responsible. Fischer could claim that because of Alex’s ongoing manipulation of his environment, Jimmy is not a person (or at least not his own person), and so cannot take responsibility for his actions. The trouble with Jimmy’s history, he might suggest, is that Alex constantly watches and intervenes to control the environment. This is like an “indirect” version of occasionalism: at each moment Alex decides to either change Jimmy’s environment, or leave it the same, and Jimmy’s action at each moment is just a causal upshot of Alex’s decision. Though Alex doesn’t directly cause each of Jimmy’s thoughts and actions, he does so indirectly via Jimmy’s environment. Being one’s own person requires not having developed under this kind of moment-to-moment control.

Taking this tack would imply what Fischer needs for his second point, as well. Fischer can clarify that NRD does not hold between marionette-manipulated agents and naturally-determined agents who perform the same actions. If Natural Jimmy is a person, but Manipulated Jimmy is not, then there is a relevant difference between the two, and it does not follow from Manipulated Jimmy’s non-responsibility that Natural Jimmy is also not responsible. Fischer can insist, though, that NRD does hold between program-manipulated agents and their natural counterparts. That is what he had in mind in discussing the theological case above. Therefore, since Natural Jimmy is responsible, a program-manipulated Jimmy would be as well. If God had created Jimmy and Jimmy’s environment to both unfold according to an internal program, so that no moment-to-moment intervention was required, then even if this program produced a sequence of events type-identical to the one’s produced by Alex’s hovering manipulation, (the program-manipulated) Jimmy would be his own person and would be responsible for his action.

I have two problems with the line of response just surveyed. First, because he

\(^{19}\)If he were to take this line of response against Pereboom’s Four Case argument, Fischer would grant that Professor Plum is not responsible because he is not a person, but would claim that the agent in case two, who is given a stable program by the neuroscientists Pereboom (2001, 113–14), is responsible.
acts on his own properly functioning biological nature, it strikes me as quite strained to say that Manipulated Jimmy is not a person (this is in contrast to professor Plum, who in my opinion is not a person for exactly the sort of reasons Frankfurt and Fischer cite). I have trouble with stretching the marionette/program distinction to cover cases of environmental manipulation.

Second, if we grant Fischer that Manipulated Jimmy is not a person, I find it mystifying that a program-manipulated Jimmy would be a person. To see why, consider this question: Why does Alex constantly tend to Jimmy’s environment? Only because he needs to, it seems. He knows that Jimmy will need to be sustained in a certain specific environment, $E$, over a certain period of time if he is to carry out the desired act. But Alex can’t maintain $E$ without a lot of intervening in Jimmy’s environment because there are many causal chains in the world not initiated by Alex, and some these tend to interfere with $E$. Alex must be constantly watching because he doesn’t know the future and so must be always ready to counteract events detrimental to maintaining $E$. It is in effect a lack of knowledge and a lack of control that drive the moment-to-moment nature of Alex’s enterprise. If he could be given more control, so that his will was effective not just in Jimmy’s immediate environment, but the whole of nature as well; and if he knew perfectly not just the laws of psychology but all the laws of nature, so that he could be sure that a single intervention in nature would eventually unfold into Jimmy’s terrorist action, then Alex could stop minding Jimmy, without ever giving up his total domination of Jimmy’s life. Intuitively, it seems that this godlike Alex would have even more control over Jimmy’s life than the more human Alex who is driven to constant watchfulness. But on the proposed account, Jimmy would be responsible for his action when controlled by the godlike Alex, because the control would not be ongoing or moment-to-moment. This is very counter-intuitive. How can increasing Alex’s level of control over Jimmy suddenly make Jimmy into his own person?

I conclude that if he is to plausibly deal with the Manipulated Jimmy case, Fischer must revise his theory of moral responsibility. Just as he reasoned with respect to the current-time-slice theories, so Fischer must now reason with respect to his own theory:

1. Manipulated Jimmy is not responsible.
2. If Fischer and Ravizza’s account were correct, then Manipulated Jimmy would be responsible.

So, (3) Fischer and Ravizza’s account is incorrect.

Fischer needs to add to his theory some further condition on moral responsibility, a condition that Manipulated Jimmy fails to satisfy (despite being a person). And if Fischer is to remain a compatibilist, this further condition must be such that Natural Jimmy does not satisfy it—it cannot be a condition that requires causal determinism to be false.

But now Fischer faces a dilemma. Either this further condition entails the absence of manipulation, or it does not. If it does not, then in principle a further manipulation case could be constructed, in which an agent satisfies this further
condition (along with all the others in Fischer’s theory), but is intuitively not responsible. For regardless of what compatibilist conditions for responsibility we specify, “there is no way to foreclose the possibility that the causes figuring in the creation of a determined morally responsible agent could be artificially fabricated.” (McKenna 2004, 176). In other words, if Fischer’s further condition does not entail the absence of manipulation, the problem will simply emerge again at another level.

If, on the other hand, the further condition does entail the absence of manipulation, then Fischer must abandon his commitment to NRD, and to hard compatibilism with it. If moral responsibility precludes manipulation, then there clearly is a relevant difference between Manipulated Jimmy and Natural Jimmy, or between any merely naturally determined agents and their intentionally determined counterparts. The difference is that the naturally determined agent satisfies all the conditions for moral responsibility, including the “no manipulation” condition, but the intentionally determined counterpart does not.

5 PROSPECTS FOR MODIFICATION

To sum up, in the previous two sections, I have argued that Fischer has two options with respect to Manipulated Jimmy. First, he can “bite the bullet” and say that Jimmy is responsible (this is the hard compatibilist line of response). Alternatively, he can say that Jimmy is not responsible, but doing this creates great pressure to abandon hard compatibilism. This is a dilemma that I think Fischer has not fully appreciated so far in his published remarks on the subject of manipulation, for his treatment of Walden Two seems to suggest a soft compatibilist line, while his treatment of theological determinism suggests a hard compatibilist position.

Whichever way he chooses to go, Fischer will need to modify (or at least better explain) his position. If he chooses to adhere firmly to his hard compatibilist line, Fischer will need to either (1) change his position and say that the residents of Walden Two are, like Jimmy, responsible or (2) explain why the cases of Walden Two and Manipulated Jimmy are different—why the mechanisms of social control in Walden Two are more problematic for moral responsibility than the mechanisms that Alex uses to control Jimmy.

If Fischer goes over to soft compatibilism, he will have to give up NRD. Suppose he says that when one person is subject to CNC manipulation by another person, even by means of otherwise “normal” or responsibility-conferring mechanisms, then the manipulated person is not responsible for her actions. This implies that there is a morally relevant difference between Manipulated Jimmy and Natural Jimmy, and more generally between any intentionally determined agent and a (merely) naturally determined counterpart. If Fischer takes this path, he must say that the relationship between moral responsibility and determinism depends on whether there is a God who uses the determinism to maintain providential control. If determinism is true and there is no God, then we can be morally responsible. But we cannot be responsible if the
deterministic world was set up by God to get us to fulfill a preestablished divine plan. In that case, we would all be victims of CNC manipulation by God.

If Fischer were to change his position on theological determinism in this way, he would be giving up what I have called his “theological neutrality intuition”: his sense that moral responsibility is neutral not only with respect to various possible truths of physics, but also with respect to various possible metaphysical truths about God. One of the main points of this paper is to show that there is serious tension between the theological neutrality intuition and our ordinary intuitive reaction to a certain type of CNC manipulation case. A plausible theory of moral responsibility can fully accommodate at most one of these intuitions.

I think Fischer should go the soft compatibilist route and reject NRD. There is not space here to fully develop a soft compatibilist version of Fischer’s theory, but in the rest of this section I would like to briefly sketch how such a development might go. In the process, I hope to show why abandoning NRD is not an obvious non-starter, and why doing so is consistent with some of (what seem to be) Fischer’s deepest intuitions.

I believe that Fischer should add a further condition to his requirements for moral responsibility, a condition implying that intentional determinism (manipulation) rules out responsibility. Abandoning NRD in this way might seem like an \textit{ad hoc} compatibilist response to manipulation counterexamples. It may seem that I am urging Fischer, when pressed into a corner, to simply pull out a “no manipulation” condition and tack it onto his theory, thereby diffusing the force of the counterexamples. I agree that this strategy would be too easy. If the further condition on responsibility is simply “no manipulation,” and the only motivation for adding it is merely to get around manipulation counterexamples, then this is indeed an \textit{ad hoc} and weak-minded response.

But things look quite different if the additional condition on responsibility is not simply “no manipulation,” but rather some distinct claim that implies the absence of manipulation, a distinct claim with some \textit{independent} intuitive motivation. If, for example, Fischer could show that some component of his current theory—plausible on its own grounds even before manipulation cases come up—implies “no manipulation,” then the soft compatibilist line of response would not be an \textit{ad hoc} maneuver, but a plausible defense against the manipulation counterexamples. This is what I think Fischer can, and should, do.

To see how this could go, recall Fischer’s account of how we come to own our volitional mechanisms as we grow up: we see ourselves as the source of our actions (and their relevant consequences), and as apt targets for praise and blame concerning them, and this view of ourselves is appropriately based on evidence. I believe that this account can be expanded in a way that captures some further aspects of ordinary developmental experience, and also provides a basis for a soft compatibilist position.

In my view, seeing ourselves as the source of our actions involves more than just seeing ourselves as their causal source. We also see ourselves as their \textit{normative} source, in a sense: the normative status of an action depends to some
extent on its connection to our reasons—on whether we performed the action intentionally, and if so on why we performed it. If another child tripped me up, it matters a great deal to my reaction whether he did so on purpose, and if so what his purpose was. He, in turn, can view himself as an apt target for my reaction only if he presumes the right sort of connection between his reasons and the physical action.

Furthermore, I think that as we develop, we normally see ourselves as the ultimate normative sources of our actions: we think that our reasons provide the most fundamental teleological or purposive explanations of our actions. We think that such explanation stops with us. This positive view of ourselves as ultimate sources has a negative consequence: we think that our reasons, as the most fundamental final causes, are not themselves merely upshots or products of some more causally basic agent’s reasons. In viewing ourselves as sources in the relevant way, we rule out that we are merely tools for some other person, whose purposes are the real final causes behind the action. This is why, as Gary Watson admits, it is “certainly vertiginous” to think of oneself as both “a full fledged free agent” and “a creature whose every move, every hope and scheme is part of another’s plan” (Watson 2000, 69).

Suppose, then, that we must see ourselves in this way—as ultimate sources of the normative status of our actions—in order to own our springs of action. On Fischer’s view, this view of ourselves must be appropriately based on evidence. As I noted earlier, this evidence requirement is best interpreted along externalist lines. This in turn implies that our belief (that we are ultimate sources) cannot be radically false. It cannot be produced in us, for example, only because our manipulator keeps herself always out of our view. Agents subject to this type of radical deception do not have beliefs appropriately based on the external evidence. So if Fischer were to accept my ultimacy condition, his evidence requirement would yield a “no hidden global manipulation” condition as a consequence.

Soft compatibilism can be seen, then, as a natural upshot of a more thorough analysis of the view we must take of ourselves in order to satisfy Fischer’s ownership condition on responsibility. I believe it can also go a long way toward accommodating what I suspect to be the fundamental intuition that lies behinds Fischer’s insistence on the neutrality of moral responsibility. Fischer’s claim that responsibility is neutral with respect to physics gains much of its plausibility from the widely shared intuition that questions about the laws of physics are “merely descriptive,” in the sense that they are concerned solely with how the world actually is, and not at all with how the world ought to be. Attributions of moral responsibility, on the other hand, are clearly normative, concerning the way in which we ought to respond to people’s actions. And changes in merely descriptive facts (e.g., determinism versus indeterminism) do not entail any changes in the normative realm. But the existence or non-existence of the traditional theistic God is not a “merely” descriptive fact: it impacts the normative realm rather drastically. And the question whether God preordains history clearly affects other normative questions besides those concerning moral responsi-
bility. For example, our view of providence will surely influence whether or not we think that there is any truly meaningless or purposeless suffering in the world, and this may affect what we consider the appropriate response to certain forms of suffering.

If what I have just argued is correct, then going over to a soft compatibilist position would not require Fischer to relinquish the core intuition behind the neutrality claim (the sense that the truth or falsehood of determinism—considered in isolation from its various possible metaphysical groundings—is a “merely descriptive” issue). Traditional theism would yield different results simply because it locates the mere fact of determinism in a metaphysical framework that imbues determinism with distinctive normative significance.20

C. P. Ragland
Saint Louis University
3800 Lindell Blvd. Suite 130
St. Louis, MO 63108 USA
E-Mail: raglandc@slu.edu

References


20 I am grateful to Kevin Timpe and John Martin Fischer for helpful correspondence on the topic of this paper, and for comments on an earlier draft.