Fischer-Style Compatibilism

John Martin Fischer’s new collection of essays, Deep Control: Essays on freewill and value (Oxford University Press, 2012), constitutes a trenchant defence of his well-known compatibilist approach to moral responsibility (Fischer 1994, 2006, Fischer and Ravizza 1998). Predominantly a collection of detailed responses to recent critics, this is not a book for beginners. It is, however, essential reading for specialists, as well as useful reading for non-specialists seeking a snapshot of the current state of the debate.

Fischer’s papers are dense with argument and alive with original and productive ideas. As is customary, however, I here focus only on lines of thought that failed to convince. In particular, I discuss two arguments: one aimed at vindicating Frankfurt-style counterexamples in deterministic contexts, and one aimed at undermining the ‘source’ incompatibilist’s conception of moral responsibility.

1. Defending Frankfurt

Once upon a time, all major parties to the freewill debate seemed to agree with the principle that an agent is morally responsible for performing a given act only if he or she could have done otherwise. The consensus ended in 1969 with Harry Frankfurt’s famous (alleged) counterexample to this Principle of Alternate Possibilities (hereafter: PAP). In Frankfurt’s case, Jones chooses and acts in such a way that we are willing to hold him morally responsible for what he does. Unbeknownst to him, however, Black was all along waiting in the wings ready to prevent him from doing otherwise (we may imagine that Black is closely monitoring Jones, ready to intervene at the first sign of his making an unwelcome decision). Yet since Black is never called upon to act, and since Jones therefore acts entirely on his own steam, most retain their judgement that Jones is morally responsible for his action. Thus he could not have done otherwise, but he is still morally responsible: PAP is refuted.
This apparent refutation of PAP is an obvious boon to those who believe that moral responsibility is compatible with causal determinism, since it suggests that we may be responsible for our actions regardless of whether determinism rules out alternative possibilities. Predictably, therefore, those who regard moral responsibility as incompatible with determinism have fought back, defending PAP from Frankfurt’s attack. Their efforts have largely coalesced around an argumentative strategy known as the Dilemma Defence (Kane 1985, Widerker 1995). This involves asking Frankfurt whether his scenario takes place in a deterministic or in an indeterministic world—either answer, it is claimed, creates insuperable difficulties for him.

To see this, suppose first that Jones acts under determinism. In that case, the incompatibilist will deny straightaway that Jones can be morally responsible for his action. So, although Jones may well be unable to do otherwise, it cannot be assumed (without begging the question against the incompatibilist) that he is responsible for what he does. So we do not have a counterexample to PAP. Suppose, second, that Jones acts under indeterminism. Now we seem to have created a problem for Black, who is unsure when to intervene. He cannot intervene before Jones has made his decision, since (given indeterminism) he has no way of knowing how Jones is going to decide before he does so. And he cannot intervene after Jones has made his decision, since then it will be too late: Jones will already have chosen from amongst open alternatives, and so will have had the power to do otherwise. It seems that there is no coherent indeterministic description of the case that serves as a counterexample to PAP. So, either way, PAP stands.

Debate rages. For the most part, proponents of Frankfurt’s counterexample have conceded that its deterministic incarnations are question-begging, and have focused their energies on formulating sophisticated versions that are able to succeed in refuting PAP on the indeterministic horn of the dilemma (e.g. Mele and Robb 1998). For his part, Fischer believes
that Frankfurt’s argument can indeed be vindicated on the indeterministic horn. But he also believes that it can be vindicated on the deterministic horn—and it is this latter claim that I wish to examine here.

In previous work, Fischer (1999 and 2002) has challenged the assumption that deterministic Frankfurt cases necessarily beg the question against the incompatibilist (see also Haji and McKenna 2004). His basic insight is this: to refute PAP, Frankfurt needs to show not that Jones is morally responsible for his action despite lacking alternatives, but rather that the issue of whether he had alternatives is irrelevant to his responsibility. So in fact the incompatibilist may be granted her assumption that, under determinism, Jones cannot be responsible for his action: the question remains whether Jones’ alleged nonresponsibility under determinism is due to his inability to do otherwise. (Note that not all incompatibilists hold that determinism rules out responsibility by ruling out alternatives (Strawson 1986, Smilansky 2000, Pereboom 2001).)

According to Fischer, the Frankfurt cases do indeed demonstrate that Jones’ lack of alternatives is irrelevant to his moral responsibility. This is because Black’s presence, which ensures that Jones cannot do otherwise, is intuitively irrelevant to the question of Jones’ responsibility. After all (so the thought goes), it would be bizarre to think that Black could make such a crucial difference just by sitting there, causally inert. So this opens the way for the compatibilist to make something like the following argument (c.f. Fischer 2012: 41-2):

1. Black’s presence renders it true that Jones cannot do otherwise.
2. Black’s presence is irrelevant to Jones’ moral responsibility.
3. Therefore, the fact that Jones cannot do otherwise is irrelevant to Jones’ moral responsibility.
4. Therefore, PAP is false.
However, the incompatibilist has what looks like an obvious retort. This is to say that it is not, in fact, Black’s presence that makes it true that Jones cannot do otherwise, but *determinism* that makes it true that Jones cannot do otherwise (Goetz 2005, Palmer 2005). If it is indeed determinism itself, and not the presence of Black, that rules out Jones’ alternatives, then (1) is false. And the argument cannot be rescued simply by replacing ‘Black’s presence’ with ‘determinism’, since it would certainly be question-begging to assume, in (2), that determinism is irrelevant to Jones’ moral responsibility. So Fischer’s argument fails.

In response to this problem, Fischer writes:

\begin{quote}
Why does the explanation in terms of causal determination ‘crowd out’ all other explanations, including the explanation in which Black plays a crucial role? Let’s say that materialism about mental states is true, and, further, that determinism obtains. So there presumably exists an explanation of an agent’s choices and behaviour entirely in terms of physical states and laws of nature. Why does it follow… that we cannot also have a perfectly good explanation of the agent’s choices and behaviour in terms of his desires, beliefs and intentions?... Why… is it *just obvious* that a prior state of the universe (together with the laws of nature) explaining why Jones cannot… choose otherwise leaves no room for *any other explanation* of Jones’ inability—such as the presence of Black, his device, and his dispositions (in a causally deterministic context)? (2012: 48)
\end{quote}

Fischer is of course correct that the argument requires only that Black’s presence be *an* explanation, and not the *only* explanation, of Jones’ lack of alternatives. Moreover, one way in which we might provide multiple (complete) explanations of the same phenomenon is by providing explanations at different *levels*. This, I take it, is how Fischer understands the situation with respect to our competing explanations of Jones’ lack of alternatives: one explanation (determinism) operates at the physical level, and the other (Black’s presence) operates at a psychological or personal level. If this were indeed the situation, then Fischer’s
argument might be sound. However, it is not: Black’s presence fails to constitute the right sort of explanation at either level, and the whole issue of competing levels of explanation is, in fact, a distraction.

Suppose that a main track runs from $A$ to $D$, with another track branching off at $B$. The set of points that would switch a train onto the branch-line at $B$ are rusted shut. So a train starting from $A$ has no choice but to continue straight on to $D$. This, I take it, is analogous to how an incompatibilist pictures a normal agent, such as Jones, facing an ‘alternative’ under determinism: determinism, like the rusted points, blocks off any genuine possibility of Jones doing otherwise.

Suppose next that the branch-line that diverges at $B$ in fact simply loops round and rejoins the main track, ending also at $D$. So it is now the case that, even were a train to switch over at $B$ (despite the rusted points), it would still end up at $D$. The looping branch-line therefore plays a role with respect to the train similar to the role that Black plays with respect to Jones (though this case is not, and is not intended to be, an exact analogy to a Frankfurt case). That is, even were Jones to begin to choose otherwise (despite determinism), Black would ensure that Jones would still end up back performing the action he is in fact determined to perform. This is then a genuine case of overdetermination. We have two perfectly good explanations of why the train will end up at $D$: because the points are rusted shut, and because the branch-line loops round. The same is true of why Jones will perform his action: because determinism is true, and because of Black’s presence. Were this the whole of the situation, Fischer would have his premise (1).

Yet matters are more complex. Recall that the reason the compatibilist is here attempting to demonstrate that Frankfurt’s counterexample refutes PAP in a deterministic context is so as to show that Frankfurt has a way through the dilemma put to him by the proponent of the Dilemma Defence. And, of course, having a way through on the
deterministic horn will be of little use to the compatibilist if it relies on her already having a way through on the indeterministic horn, for then the Dilemma Defence has already been put to rest. So in evaluating attempts to save Frankfurt’s argument on the deterministic horn, we should assume that we do not already have a way of saving his argument on the indeterministic horn. That is, we should assume that, absent determinism, Black’s strategy would not succeed in fully expunging Jones’ alternatives; Black needs determinism. Otherwise the compatibilist is simply expending philosophical effort for no dialectical gain.\footnote{Fischer (2011: 43-4) mentions two other reasons why it might be useful for a compatibilist to formulate a successful deterministic version of a Frankfurt case, but neither undermines this point. The first is that some think that determinism is necessary for moral responsibility. Yet these philosophers are not precluded from using indeterministic Frankfurt cases to impugn PAP, since they may adopt their opponents’ assumption that Jones may be morally responsible under indeterminism simply for the sake of proving that open alternatives are unnecessary for responsibility. The second is that it may be easier to construct a successful Frankfurt case given determinism than given indeterminism. Yet if a deterministic version succeeds where the parallel indeterministic version fails then determinism plays a crucial role, and we are (again) correct to assume that determinism is required in order for Black to do his job.}

Perhaps it will be objected that, in order to have a way through the dilemma on the indeterministic horn, we need a case \textit{both} in which Black succeeds in expunging Jones’ alternatives \textit{and} in which Jones is morally responsible for his action. So we could have a case in which, absent determinism, Black’s strategy would succeed in expunging alternatives, but in which we nevertheless do not have a way through on the indeterministic horn, since it cannot be shown that Jones remains responsible. Yet this would therefore be a case in which Black expunges Jones’ alternatives \textit{in a way that also expunges his moral responsibility}. If Black’s intervention has this effect under indeterminism, there will be reason to think that it will have this effect under determinism, and so (2)—the premise concerning the irrelevance of Black’s presence to Jones’ responsibility—will be undermined. In attempting to produce a
deterministic Frankfurt case, therefore, it is simply of no dialectical use to the compatibilist to assume that Black can pull off his trick without the help of determinism.²

So suppose then that there is a secondary branch-line branching off from the first branch-line at C (see Figure 1). This secondary branch-line does not loop back round to D. Yet, as with the turning at B, the points at C are rusted shut. So a train taking the first branch-line will, as before, end up at D, but this is now the case only because the points at C are rusted shut, and not because of the nature of the branch-line itself. In sum: a train, starting from A, will end up at D; it will do so, first, because the points at B are rusted shut and so it cannot turn off from the main-line; second, because even if it were to turn off at B, it would proceed back round to D, due to the nature of the branch-line together with the fact that the points at C are also rusted shut.

Say that we now ask the question: why does a train, starting from A, have no alternative but to end up at D? Is it (i) because the points are rusted shut at B and C, or (ii) because of the partially looping first branch-line? I take it that the answer is clearly (i).

²This is a dialectical point that may not have been fully appreciated by Haji and McKenna (2006). In support of (1)—the claim that Jones’ lack of alternatives is due to Black—they introduce a case, taken from Pereboom (2003), that purports to show that Black can eliminate Jones’ alternatives (in a responsibility-preserving way) under indeterminism. It is indeed true that, if this case stands up to scrutiny, we have reason for accepting (1). But it is also true that, if it stands up to scrutiny, it already provides a vindication of Frankfurt’s argument (under indeterminism), thus rendering redundant the attempt to vindicate it under determinism.
Moreover, parallel conclusions hold for Black and determinism, and they do so regardless of the level at which we describe Frankfurt’s case. Say that we describe it entirely mechanistically, considering only Jones’ physical states, and perhaps replacing Black with a simple robot that monitors Jones’ neurological processes and is primed to intervene in them under certain conditions. Thus Jones will end up performing his specified action come what may. This is because, first, determinism ensures that his neurological processes can unfold in only one way; second, because even were they to begin to unfold in some other way, they would still ultimately issue in the same result, due to the robot’s presence together with determinism. I take it to be intuitively obvious that the explanation of Jones’ lack of alternatives is here the truth of determinism, and not the presence of the robot (just as the explanation of the train’s lack of alternative destinations is the rusted points, and not the partially looping branch-line).

Now say that we adopt instead the intentional stance (Dennett 1973), and describe the case with reference to Jones’ and Black’s psychological states and processes. Note that, in order to maintain dialectical relevance, we must assume that causal determinism holds at this level of explanation—since Black must be able to know, on the basis of Jones’ psychological states, what Jones will decide in advance of his deciding it. That is, if we had a method of making the argument work at the psychological level without assuming psychological determinism—that is, under psychological indeterminism—then we would already have a method of answering the Dilemma Defence on the indeterministic horn, and we would not need this argument in the first place. So Jones has no alternative but to act as he does: this is due, first, to the truth of psychological determinism, and, second, to the presence of Black together with the truth of psychological determinism. The shift to the intentional stance makes no difference to the structure of the argument. It is, again, intuitively obvious that it is here the
truth of psychological determinism, and not the mere presence of Black, that explains Jones’ lack of alternatives.

I am not convinced, therefore, that Fischer has shown there to be a sound response to the Dilemma Defence on the deterministic horn of the dilemma. Yet Fischer also indicates a backup position (2012: 42-7; see also Haji and McKenna 2004). This is that, even if deterministic versions of the Frankfurt cases turn out to beg the question against the committed incompatibilist, they need not do so against the ‘floating voter’ not yet persuaded by either party to the debate. In particular, Fischer considers an audience that is as yet unsure whether or not determinism rules out alternative possibilities. To this audience, a Frankfurt-style compatibilist might argue as follows (c.f. Fischer 2012: 44-6):

(1) It is uncertain whether determinism alone prevents Jones from doing otherwise.
(2) It is certain that Black (together with determinism) prevents Jones from doing otherwise.
(3) Intuitively, Black’s presence is irrelevant to Jones’ moral responsibility.
(4) Therefore, the fact that Jones cannot do otherwise is irrelevant to Jones’ moral responsibility.

However, it is not immediately obvious that there is indeed a substantial constituency to which this particular argument can appeal. The problem is with (1). Any potential audience for this argument must at least understand what the doctrine of determinism is. They will, therefore, understand that determinism rules out alternative possibilities in some senses and not in others, and they will also understand, presumably, just which sorts of alternatives it rules out and which it does not. After all, this is not an issue that is typically treated as controversial. Nearly everyone agrees, for instance, that determinism does not take away merely conditional possibilities (e.g. of the form ‘I would if I tried’), or counterfactual possibilities where the past
and the laws are not fixed, though it does take away counterfactual possibilities where the past and the laws are fixed, as well as any kind of physical possibility. What is controversial, and forms the major focus of the dispute, is which (if any) of these types of alternative are relevant to moral responsibility: when philosophers argue over whether determinism rules out alternatives, they are arguing over whether determinism rules out alternatives in the sense required for moral responsibility.

To defeat this argument, then, it seems that all an incompatibilist need do is separate out the various possible senses in which Jones might be ‘prevented from doing otherwise’, and invite her opponent to run separate versions of it corresponding to each of these senses. As regards any one version of it there will not, I expect, be much uncertainty whether determinism prevents Jones from doing otherwise in that particular sense (so long as the audience meets the minimal condition of understanding what is meant by ‘determinism’). Since the audience will in no case be disposed to accept (1), each version of the argument, and hence the general version of the argument set out above, can be expected to fail.

2. ‘Source’ Incompatibilism

While most incompatibilists have traditionally held that determinism rules out moral responsibility by virtue of eliminating alternatives, ‘source’ incompatibilists claim that determinism rules out responsibility directly, by eliminating us as the ultimate sources of our actions. The idea is that to be truly responsible for an event the buck for producing that event must stop with us, but that, under determinism, the buck never stops with us, always being passed on back into the indefinite past. Thus ‘we cannot shirk the perspective from which all that happens is ultimately a matter of luck’ (Smilansky 2000: 284-5).

We might understand the basic source-incompatibilist argument as follows:
(1) It is fair to hold a person morally responsible for her actions only if she is the ultimate source of those actions.

(2) If determinism is true then no person is the ultimate source of her actions.³

(3) Therefore, it is never fair to hold a person morally responsible for her actions.

Fischer attempts to undermine this argument by challenging (1), arguing that it encapsulates an ‘over-inflated’ and ‘metaphysically extravagant’ conception of moral responsibility. According to him, the source incompatibilist holds that moral responsibility for an action requires total control over that action, where ‘An agent has total control over X if and only if for any factor f which is a causal contributor to X and which is such that were f not to occur, then X would not occur, the agent has control over f’ (2012: 171). Yet this, Fischer argues, is an implausible requirement, since not only do we clearly lack such total control over our actions, but we lack it in ways that do not, intuitively, seem to have much bearing on our moral responsibility.

As Fischer points out (2012: 168-71), there are indefinitely many facts that are causally necessary for our performing the actions we do: for instance, that the sun continues to shine, that the air continues to be breathable, that we have not just been struck by lightning or a meteorite, that there has not just been an earthquake, that our parents fed and cared for us, and so on. We are not responsible for any of these things—but, equally, our nonresponsibility for these causal factors seems clearly irrelevant to the issue of whether we are morally responsible for what we do. No one was ever excused of a crime on the grounds that she lacked control over whether or not she was stuck by lightning just prior to acting: our ordinary notion of moral responsibility allows plenty of room for this kind of luck.

Smilansky has responded to this challenge as follows:

³ And perhaps even if it is not true: see for instance Strawson 1986 and Pereboom 2001.
Luck is undoubtedly present in our lives, but the central question is how its presence is manifested… luck is not located in some corner but, when we look deeply, we see that it goes *all the way through*… Fischer attempts to turn the table on the sceptics by pushing them to acknowledge that they themselves do not mind the presence of a large measure of luck. But this wrongly locates the sceptical worry, which is whether, if we look deeply, anything but luck *remains.* (2003: 274-5)

Fischer replies (2012: 176-7) by first conceding that, while his examples all concern uncontrolled *necessary* conditions of our acting as we do, determinism seems to entail the existence of an uncontrolled *sufficient* condition of our acting as we do (i.e. the past together with the laws of nature)—and that lacking control over a sufficient condition may well be a different matter from lacking control over a merely necessary condition. Nevertheless, Fischer issues the following challenge to source incompatibilists. Inviting us to consider Figure 2, he writes:

My question is this: if one is not troubled by the existence of the vertical line, why be troubled by the horizontal line? They are both the same in the sense that they represent ‘external’ factors that are entirely outside the relevant agent’s control; in virtue of what is the horizontal line troubling in a way in which the vertical line isn’t? A mere appeal to ‘externality’ will not distinguish the two lines—they are equally ‘external’ to the Agency Line. Similarly, the sun is ‘external’ to the agent in just the same way as the antecedent causal sequence—each equally impugns Total Control, and both introduce just the same sort of luck. (2012: 183)
Yet we must be careful to distinguish *totality* from *ultimacy*. Whereas the totality of one’s control may be understood along the lines Fischer outlines, the ultimacy of one’s control is surely not to be understood in terms of the *number* of causal factors one controls but in terms of the *source* and *nature* of one’s control over them. Specifically, the crucial issue for ultimacy is whether the relevant causal chains terminate in one’s own agency or carry on through one’s agency and onwards into the distant past. Indeed, ultimacy in this sense is neither necessary nor sufficient for totality. Suppose that one has total control over the occurrence of some event, in Fischer’s sense of having control over every necessary condition of its occurrence. If one controls these causal factors merely in the way that a thermostat controls the temperature of a room, then one’s control is not ultimate in the relevant sense, despite its totality. Conversely, if one has ultimate control over just one necessary condition of an event’s occurrence, then one has a kind of ultimate control over its occurrence (in the sense that one yields a ‘veto’ over it), even if this control is merely partial.

To elaborate: suppose that you are preparing to perform some action $A$. And suppose that you have ultimate control over only one of the many causally necessary sustaining conditions mentioned by Fischer, such as the condition that you not be struck by lightning. That is, we are imagining a bizarre case in which you have ultimate control over lightning, but not over anything else—including the beliefs and desires that are deterministically (let us suppose) now leading you to do $A$. Since not being struck by lightning is a necessary condition of your doing $A$, your ultimate control over its striving grants you a sort of veto over whether you do $A$, and therefore ultimate (though partial) control over your doing $A$. It puts the issue of whether or not you do $A$ in your own hands, albeit crudely. Similarly, I take it, such ultimate though partial *control* over your doing $A$ also renders you ultimately though partially *responsible* for doing $A$. (Of course, since your only available means of preventing yourself from doing $A$ are so drastic, we are unlikely to *blame* you for doing it—unless $A$ is something
very heinous indeed, in which case we may well blame you for not striking yourself down with lightning.)

Thus the source-incompatibilist, having distinguished ultimacy from totality, may reasonably claim that ultimate control over any causal contributor to an event is sufficient to get moral responsibility for that event off the ground (Istvan 2011). The totality of one’s control—and hence the totality of one’s responsibility—is then a subsequent issue. The source incompatibilist may therefore agree with Fischer that there is no relevant difference between the horizontal and vertical lines in Figure 2, insisting simply that were the agent to be the ultimate source of either line, she would then bear a kind of ultimate responsibility for what she does. And while we generally take it for granted that we are not the ultimate sources of the factors represented by the vertical line, determinism guarantees that we are not the ultimate sources of any of the factors represented by either line. This, then, is how determinism might be taken to rule out moral responsibility on source incompatibilist grounds.

Of course, the compatibilist will simply deny that responsibility requires any amount of such ‘ultimate’ control: mere thermostat-style control, or some highly sophisticated version of this, may be claimed to be sufficient. So it is unlikely that the incompatibilist’s invocation of the sourcehood requirement will grant her an argument capable of swaying a committed compatibilist. However, this was presumably never the source incompatibilist’s aim. Instead, her talk of sourcehood and ultimacy is best heard as an attempt to articulate a central incompatibilist concern in a way that makes it compelling to ‘floating voters’ not yet under compatibilism’s sway. This, as we have seen, is a perfectly reasonable dialectical aim—and it is not one, at least as far as I can see, that need be undermined by Fischer’s line of attack.
References


