

Moral Responsibility and Alternative Possibilities

Essays on the Importance of Alternative Possibilities

Edited by
David Widerker
and
Michael McKenna

ASHGATE

Chapter 4

In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing

Carl Ginet

It is very natural to think that a person deserves blame or credit for a certain thing's being the case, is morally responsible for it, only if she could have made it otherwise. I deserve blame for the fact that the car is not yet repaired only if I could have made it the case that the car was repaired by now. Smith deserves credit for knocking down the opposing quarterback only if he could have avoided knocking down the quarterback. It's natural to embrace the following general principle, to which I will give the name 'the principle of alternative possibilities' (PAP for short):

PAP: An agent *S* is morally responsible for its being the case that *p* only if *S* could have made it not the case that *p*.

A slightly stronger principle is equally plausible, namely, that *S* is morally responsible for it's being the case that *p* only if *S* could have made it not the case that *p* by some means that *S* knew about, or should have known about, in time to do so. The driver whose cup of coffee was, unbeknownst to her, laced with a reaction-slowing drug does not deserve to be reprehended for its being the case that while subsequently driving her car she reacted too slowly to avoid hitting a pedestrian, because there is no way she could have prevented that's being the case that she knew about in time to do so; she could have prevented it by not drinking that coffee or by not driving after drinking it but it's not the case that she knew or should have known this. But the ordinary drunken driver may be culpable for his reacting too slowly to avoid hitting a pedestrian, because he knew or should have known that he could have prevented such a situation by not drinking before driving or not driving after drinking. Hereafter I will take this strengthening of PAP as understood. That is, I will read '*S* could have made it not the case that *p*' as elliptical for '*S* could have made it not the case that *p* by some means *S* knew about or should have known about in time to do so.'

PAP was, I think, widely taken for granted until twenty-five years ago when Harry Frankfurt challenged it in his well-known paper titled 'Alternate

Jones then knew or should have known how he could prevent it. For the only way Jones could have prevented it would have been by preventing Black's mechanism from being set up or by disabling it after it was set up. But there was no time at which Jones knew or should have known anything about the existence of Black's mechanism.

Now for the crucial question. Is Jones morally responsible for its being the case that he does *B* by *t3*, despite the fact that he could not avoid the obtaining of that state of affairs? It is Frankfurt's intuition that he is responsible for it, since Jones does *B* by *t3* on his own without the intervention of the mechanism and would have done *B* by *t3* in the same way for the same reasons even if the mechanism had not been there and he could have avoided doing *B* by *t3*. I do not share that intuition. I do, however, have the intuition that there is something here for which Jones may be responsible, something that entails his doing *B* by *t3* – namely, his doing *B* at the precise time he actually did it, *t1*. This latter state of affairs is one whose obtaining Jones could have avoided (though, if he had avoided it, he would have done *B* at some later time between *t2* and *t3*). Jones's doing *B* at *t1* differs from Jones's doing *B* by *t3*, only in being temporally more specific. My intuition is that Jones is not responsible for the obtaining of the temporally less specific state of affairs, because, owing to the presence of Black's mechanism, Jones could not have avoided it, but he may be responsible for the obtaining of the temporally more specific state of affairs, which he could have avoided. This phenomenon of being responsible for the obtaining of a more specific state of affairs while *not* being responsible for the obtaining of an entailed less specific state of affairs is not at all uncommon. I am, for example, responsible for my being now in the particular room I'm in but I am not responsible for my being now within a one-million-mile radius of the center of the earth. I'm inclined to think that the attraction of Frankfurt's intuition about his example arises from failure to distinguish the two states of affairs, Jones's doing *B* at *t1* and Jones's doing *B* by *t3*, which differ only in their temporal specificity.

This failure to distinguish is made all the easier by the fact that the more specific, Jones's doing *B* at *t1*, is morally equivalent to the less specific, Jones's doing *B* by *t3*, in the sense that the first is as blameworthy, and for the same reasons, as *would be* the second were Black's mechanism not present. This moral equivalence between Jones's doing *B* at *t1* in the actual case and Jones's doing *B* by *t3* in the case where Jones is responsible for it helps one to be careless about the distinction between them in the actual case, encourages the feeling that it doesn't much matter whether we get exactly right – with a precision that would satisfy a philosopher – just what Jones is responsible for here.

If my intuitions about the example do not initially seem compelling they may become more nearly so if you look at the matter from Black's point of view. Ever since he set up his mechanism Black knew that in one way or another Jones would do *B* by *t3* and he will surely find it natural to deny that Jones is responsible for that fact: after all, Black has arranged things to ensure it, has set up a condition that is, in the circumstances, sufficient for it. And Black will also find it natural to say that, if it turns out that Jones does *B* at some time before the deadline *t2*, then Jones *will* be responsible for *that* fact.¹

Notice that I'm not saying that what Jones may be responsible for is its being the case that he does *B* on his own rather than as a result of Black's mechanism. It would be a mistake to say this. Since Jones was completely unaware of Black's mechanism, he neither knew nor should have known the fact expressible by his saying, 'I am doing *B* on my own rather than as a result of Black's mechanism'; but he could have been responsible for making that fact obtain only if, at the time, he knew or should have known that he was doing so. What I say Jones may be responsible for is its being the case that he does *B* at *t1*. He did know that he was making that fact obtain. (Must he have known that the time at which he was doing *B* was *t1*? Yes, he must then have been aware of that time, *t1*, under the description 'now'.)²

So my intuition about Frankfurt's example, again, is that, although Jones may be responsible for its being the case that he does *B* at the precise time *t1*, something he could have avoided, he is not responsible for the obtaining of the temporally less specific state of affairs of his doing *B* by *t3*, because, given the presence of Black's mechanism, that is something he could not have avoided. This finding as to what Jones is and is not responsible for depends on there being a difference between the time at which Jones actually does *B* and the time at which Jones would do *B* if he were caused to do it by Black's mechanism. A defender of Frankfurt's argument might think that my finding could be blocked by revising the example so that this difference is eliminated, so that the precise time at which Jones actually does *B* and the precise time at which Black's mechanism would cause Jones to do *B* are the same.

But I think that such a move cannot succeed in saving Frankfurt's argument. For in order to ensure that the two times are the same Black's mechanism would have to be set up so that it would be triggered by the absence of some condition whose actual presence is causally sufficient to ensure that Jones does *B* at *t1*. That is, it would have to be posited, first, that there occurred at a time *t0* prior to Jones's doing *B* at *t1* a condition *C* that was causally sufficient for his doing *B* at *t1* and, second, that Black's mechanism was rigged so that, had *C* failed to occur at *t0*, the mechanism would have causally necessitated Jones's doing *B* at *t1*.³ Condition *C* at *t0* must be such that, its obtaining makes it the case that it is not open to Jones after *t0* to avoid doing *B* at *t1*, that is, from *t0* on Jones could not avoid doing *B* at *t1*. Otherwise, if there were no such condition *C* whose presence at *t0* forces Jones's doing *B* at *t1* and whose absence at *t0* would cause Black's mechanism to force Jones's doing *B* at *t1*, then there would be nothing in the example that entails that Jones could not avoid doing *B* at *t1*.

Now what should we say about this modified example? What can Jones be responsible for here? That depends, it seems to me, on whether or not the occurrence of *C* at *t0* is such that the following three things are true of it: (a) Jones could have prevented it; (b) he knew or should have known how he could do this; and (c) he knew or should have known that it would mean that from *t0* on he would be unable to avoid doing *B* at *t1*. If the occurrence of *C* at *t0* does satisfy these conditions, then we PAP-sympathizers will find it natural to say that what Jones may be directly responsible for here is his not trying to avoid his doing *B* at *t1* by preventing, or not contributing, the occurrence of *C* at *t0*,

his allowing the occurrence of *C* at *t*0 to make him unable to avoid doing *B* at *t*1. And his allowing this is morally equivalent to his doing *B* at *t*1 in the circumstance where he could up until *t*1 have avoided that, because it is not the case that he knows or should know anything about Black's mechanism and so he is justified in supposing that he could avoid doing *B* at *t*1 by avoiding the occurrence of *C* at *t*0; he has no reason to think otherwise.

What if, on the other hand, the occurrence of *C* at *t*0 is *not* such that (a) Jones could have prevented it, (b) he knew or should have known how he could do this, and (c) he knew or should have known that its occurrence would force his doing *B* at *t*1? Then it is no longer true of the example that it is only the presence of Black's mechanism that makes it the case that Jones could not avoid doing *B* at *t*1. For, even if Black's mechanism were not there, it would still be true that Jones could not avoid doing *B* at *t*1. This follows from the supposition that one of (a), (b), or (c) does not hold. If (a) does not hold, then Jones could not prevent the occurrence of *C* at *t*0. From this and the fact that he could not avoid doing *B* at *t*1 once *C* did occur at *t*0 it follows that he could not ever avoid doing *B* at *t*1. If (b) or (c) does not hold, then either Jones did not know how to prevent the occurrence of *C* at *t*0 or did not know that if *C* occurred at *t*0 he would then be unable to avoid doing *B* at *t*1; and so, whichever is the case, he did not know how to avoid its being the case that he did *B* at *t*1 and thus, in our enriched sense of the expression, could not avoid its being the case.

Now, if the example is such that, even apart from the presence of Black's backup enforcing mechanism, Jones could not avoid its being the case that he did *B* at *t*1, then the strategy of Frankfurt's argument is undermined. For the strategy was to take an action that satisfies PAP and add a factor, the backup enforcing mechanism, that makes the action no longer satisfy PAP and then argue that, since it is only the presence of this merely counterfactually intervening mechanism that makes the action unavoidable and Jones would have done the same thing for the same reasons even if the added factor were not there and he could have done otherwise, Jones is still responsible for his action.

The feature of the example that is the basis for Frankfurt's intuition about it is this feature that, since it is only the presence of this merely counterfactually intervening mechanism that makes it the case that Jones could not have done otherwise, Jones's performing the action would have obtained for the same reasons even if he could have done otherwise. Now – in the case where not all of (a), (b), and (c) hold – this feature is gone. Now the presence of the merely counterfactually intervening mechanism does no work. For now we are asked to accept that had the mechanism not been added to the example Jones would have been responsible even though then too he could not have done otherwise. But this is just to assume what was to be proved. The argument that was to persuade us to give up PAP can no longer be that the addition of the counterfactually intervening mechanism gives us no reason to change the judgment that Jones is responsible that we were willing to make about the original case, before the mechanism was added to the example and Jones could have done otherwise. For now in the original case Jones could *not* have done otherwise and the judgment that we were happy to make in the other original

case (the original original case) now implies the falsity of PAP: so naturally we who start out as PAP-sympathizers are not willing to make that judgment about this new original case until we are persuaded by some argument that we should do so. Frankfurt's argument can take it as a premise that it is the appropriate judgment to make only at the cost of begging the question.

To sum up: my earlier finding about Frankfurt's example cannot be successfully evaded by revising the example so as to make the time at which Jones would do *B* were Black's mechanism to intervene coincide precisely with the time at which Jones actually does *B*. Such a revision either makes the argument beg the question or permits a PAP-sympathizer to find that what Jones may be responsible for is, not his doing *B* at *t*1, but his allowing or contributing the occurrence of a prior condition that he knows will make his doing *B* at *t*1 unavoidable.

The central premise of Frankfurt's argument, which he takes to be illustrated by his example, is enunciated in the following remarks (from the first two paragraphs of section V of Frankfurt, 1969):

Now if someone had no alternative to performing a certain action but did not perform it *because* [my emphasis] he was unable to do otherwise, then he would have performed exactly the same action even if he *could* have done otherwise. The circumstances that made it impossible for him to do otherwise could have been subtracted from the situation without affecting what happened or why it happened in any way. Whatever it was that actually led the person to do what he did, or that made him do it, would have led him to do it or made him do it even if it had been possible for him to do something else instead ... When a fact is in this way irrelevant to the problem of accounting for a person's action it seems quite gratuitous to assign it any weight in the assessment of his moral responsibility. (p. 837)

Let us restrict attention from now on to actions that satisfy whatever conditions other than ability to do otherwise are necessary for responsibility (conditions having to do with the agent's knowledge or intention about what she is doing). About such cases Frankfurt commits himself to the following proposition:

S is morally responsible for acting as she did, despite the fact that *S* could not have acted otherwise, if (a) it's not the case that she acted as she did because she could not have acted otherwise and (b) she would have acted in the same way for the same reasons even if she could have acted otherwise.

Rephrased in the terminology of action states of affairs, this becomes what I will call thesis F:

F: *S* is morally responsible for its being the case that she acted in a certain way at a certain time, despite the fact that she could not have avoided its being the case, if (a) it's false that it was the case because she could not have avoided its being the case and (b) it would have been the case for the same reasons even if she could have avoided its being the case.

It may not be obvious at first glance how to interpret condition (a), that it's false that it was the case because she could not have avoided its being the case. I take it to mean simply that there are facts that sufficiently explain *S*'s acting as she did that do not include the facts that make it the case that she could not do otherwise. So understood, condition (a) is satisfied whenever condition (b) is. If *S* would have acted in the same way for the same reasons even if she could have done otherwise, then *S*'s action is the outcome of antecedent events and circumstances that would still be there to explain it if there did not obtain the facts that make it the case that she could not do otherwise; and if they would explain it in that counterfactual situation, they explain it in the actual situation.

As I've indicated, I do not find thesis *F* intuitive, even after considering the example. I wish I could decisively refute the thesis by presenting a clear, "open and shut" counterexample to it. But the counterexamples I have been able to think of are likely to be contested by those who are inclined to share Frankfurt's intuition about his example. The best I can do is to try to cast doubt on thesis *F* by offering clear counterexamples to the generalization of the thesis to all states of affairs, and suggesting that, in the absence of any good reason to think that the specific thesis should not be just as false as the general one, we have reason to doubt the specific thesis. And if we who are disinclined to share Frankfurt's intuition about his example have reason to doubt *F*, then we have reason to reject the only argument we've been offered that we *should* share Frankfurt's intuition.

Here's the generalization of the thesis (call it *G*):

G: *S* is morally responsible for its being the case that *p*, despite the fact that she could not have avoided its being the case that *p*, if (a) it's false that it was the case because she could not have avoided its being the case and (b) it would have been the case for the same reasons even if she could have avoided its being the case.

It's easy to think up clear counterexamples to thesis *G* where the proposition *p* says that *S* did *not* act in a certain way.

Example (1): Smith is a member of an organization whose sole aim is to bring about the abolition of the monarchy. She has been invited to attend a reception for the queen. The custom at these affairs is that when the queen enters the room everyone else stands up and applauds. In conformity with the wishes of her anti-monarchist group, Smith is resolved to show contempt for the monarchy by not standing up when the queen enters. But other members of the group are concerned that Smith will not stick with her resolution and so they have attached to Smith's chair a device which will, unnoticed by Smith and shortly before the queen enters, put across Smith's lap a strap that will render her unable to stand up when the queen enters even if she tries to. (Or we can imagine that the device temporarily paralyzes Smith.) But Smith never even considers abandoning her intention and when the queen enters she intentionally does not stand. It seems that on thesis *G* Smith is responsible (and blameworthy or praiseworthy, depending on your point of view) for its

being the case that she does not stand up when the queen enters. For, even if Smith could have stood up, she would have remained firm in her intention not to do so for the same reasons she actually decided not to do so, and those reasons included no belief that she could not stand up. Yet it is abundantly clear that, because there was a physical barrier that made it impossible for Smith to stand up, it would be wrong to blame or praise her for not actually standing up. Of course, Smith may still be to blame or to praise for not *trying* to stand up, which is something she could have done, or for *intentionally* not standing up, which is something she could have avoided.⁴

This is a counterexample to thesis *G*, for (a) and (b) are satisfied with respect to its being the case that Smith does not stand up. This is because we stipulated in the example that, even if she could have stood up, Smith would have stuck with her decision not to do so for the same reasons that she actually decided not to do so. So there are facts sufficient to explain her not standing up – those that explain her not trying to stand up – that do not include the facts about her being strapped to the chair which made it the case that she could not have stood up. They are sufficient to explain it because they would explain it in the possible world where she could have stood up.

Example (2): Now imagine the following change in the preceding example. The device that Smith's fellow anti-monarchists have installed on her chair is able, not only to put in place a strap that will hold her down, but also to monitor her neural processes and make erecting the barrier contingent on certain developments it detects there. The device is rigged so that the barrier will be instantly put in place if either Smith's brain issues motor impulses that could be the initial stages of an effort to rise or Smith makes a decision to rise. But it never actually puts the barrier in place because, as in the preceding example, Smith never considers rising and never engages in any even incipient attempt to rise but remains firm in her intention to stay seated as the queen enters. Again the device attached to Smith's chair makes it the case that her standing up when the queen enters is not open to her and therefore, it seems to me, makes it the case that Smith is not morally responsible for not standing. The changes in the way the device renders Smith unable to stand – not through placing a physical barrier to her standing but through being ready to erect such a barrier should Smith attempt or decide to stand – are no reason to change that judgment.

Example (3): Now imagine one more change in the device attached to Smith's chair. As before it monitors her neural processes but now, should she decide to rise or begin to try to rise, the device, instead of placing a physical barrier to her rising, manipulates her brain so as to cause her to reconsider and to decide not to rise and to remain firm in that intention until after the queen is in and seated. (These changes in the way the device renders Smith unable to stand up bring the example quite close to simulating, for a case of not acting, Frankfurt-style examples for cases of acting.) Do these changes give us any reason to alter our judgment that Smith is not responsible for not standing up? I cannot see how it could be plausibly argued that they do. It still seems clear that, because

the action of standing up is not open to Smith, she cannot be held responsible for not performing that action.⁵

Such examples show, I think, that the generalization G of Frankfurt's thesis F fails to hold for these cases of *not* acting in a certain way (and they suggest that PAP does hold for such cases). Of course, one could grant that thesis G fails to hold and with logical consistency maintain that the specification of G to cases of *acting* in a certain way, Frankfurt's thesis F, does hold (and that PAP specified to such cases does not hold). This position has in fact been taken by John Fischer (1986).

About examples of non-actions like (1)–(3) above, where the agent does not try to act in a certain way but in any case could not have acted in that way, Fischer shares my intuition that the agent cannot be held responsible for not acting in that way (but at most for not trying to do so); but about Frankfurt-style examples of actions, he subscribes to Frankfurt's intuition that the agent is responsible for acting in a certain way despite the fact that she could not have avoided acting in that way. He tries to explain his asymmetrical intuitions by invoking the idea that what is required for responsibility for an action or non-action is, not ability to act otherwise, but *control*. He takes it that an action, doing *X*, entails causing a certain upshot *X'*. He proposes that one is responsible for doing *X* or for not doing *X*, as the case may be, only if one has control over this event *X'* that is constituent in doing *X*. And he premises that one has the requisite control over the event if one does *X* by causing *X'* in a 'suitable way.' Of course, if what one is responsible for is *not* doing *X*, then one has not caused the constituent event at all and one's control over it in that case has to consist in one's having had it in one's power to cause it.

That is the gist of Fischer's explanation. Its defect is that it depends on the premise that causing the constituent event *X'* in a suitable way is sufficient to make one responsible for doing *X* (provided, of course, that necessary conditions independent of control or ability to do otherwise are satisfied, conditions such as one's knowledge or intention about the action). To see the difficulty with this premise, consider again the first version of our Black/Jones example and suppose that things go differently than we stipulated before. Suppose that at *t*1 Jones decides *not* to do *B* and resolutely sticks to this intention until the deadline *t*2, whereupon Black's mechanism is triggered and forces Jones to decide to do and to do *B* by *t*3 and it does this by causally necessitating a sequence of events that constitutes Jones's causing the constituent event of doing *B* (which is, presumably, the death of the aged relative) in a suitable way. I doubt that even Frankfurt would want to say that Jones was responsible for his doing *B* by *t*3 if it came about in that way.⁶

So Fischer has not given a satisfactory explanation of why thesis G should hold for cases of action but not for the cases of non-action I described. And as long as I – who do not share Frankfurt's intuition and offer reason to suspect that its basis is either a failure to distinguish a more from a less specific state of affairs or else an argument that begs the question – see no explanation of why there should be a difference in truth-values here, it is reasonable for me to take the clear falsity of the general thesis G as reason to doubt the truth of its

specification to actions, thesis F. And if F fails to hold, then Frankfurt's argument for his intuition collapses. So it's reasonable for me to doubt that Frankfurt has given us good reason to doubt PAP.⁷

Addendum (2001)

The argument of the first half of this chapter (to the paragraph beginning 'To sum up, on page 81') presents the constructor of a Frankfurt-type example (an FTE) with a dilemma: either the added fail-safe mechanism allows it to be the case that *S* could have avoided acting in the desired way at the precise time she did so; or its making unavoidable that action event at that precise time depends on the event's being deterministically caused by prior conditions in the world, in which case the PAP-sympathizer can hardly be expected to agree that *S* was morally responsible for that event before the fail-safe mechanism was added.

Alfred Mele and David Robb claim to have found a way for an FTE-structor to avoid this dilemma. They start with an example in which the desired action is *S*'s *deciding* at a certain precise time *t* to act in a certain way *A*, and in which this decision event is *indeterministically* caused, so that initially the example satisfies PAP. (Because the causation was indeterministic it does not entail that *S* could not have avoided deciding at *t* to *A*.) Then they add to the example that there actually occurred right up to *t* a process that would have deterministically caused *S*'s deciding at *t* to *A* but for the fact that it was preempted by the other process that indeterministically caused that event. The upshot, they claim, is that we have an example where *S*'s deciding at *t* to *A* was something *S* could not have avoided (because of the presence of the deterministic process) but also something that was not deterministically caused (because the deterministic process was preempted by an indeterministic one); and since *S*'s deciding at *t* to *A* was not deterministically caused, there is no reason to hesitate to judge that *S* was responsible for her deciding at *t* to *A*. (Some PAP-sympathizers may be inclined to think that an agent cannot be responsible for an action that is indeterministically caused. I am not inclined to think this but, in any case, there is another problem with Mele's and Robb's argument.)

I doubt that the Mele–Robb example is coherent. Let me explain why. Let us give the label *E* to the event of *S*'s deciding to *A*. The short interval of time *t* occupied by the event *E* began, let us say, at instant *t*1 and ended at instant *t*2; that is, *t* was the interval [*t*1–*t*2]. The final stages of both the deterministic process and the indeterministic process occur during an interval that is before and continuous with [*t*1–*t*2], say [*t*0–*t*1]. Let us name those final stages of the deterministic and indeterministic processes *D*-final and *I*-final, respectively. What I doubt is that the following three claims that Mele–Robb make about the example can all be true:

- (1) *I*-final preempts *D*-final as the immediate (most proximate) cause of *E* at [*t*1–*t*2].

- (2) If *E* had not been indeterministically caused by *I*-final, then *D*-final would have deterministically caused *E* to occur at precisely the same time as *I*-final caused it to occur, namely, precisely in the interval $[t1-t2]$.
- (3) The occurrence of *E* at $[t1-t2]$ was unavoidable by *S*.

If (1) is true then one or the other of the following two things must be true:

- (1a) *I*-final's causing *E* (and nothing short of that) is what blocked *D*-final from causing *E*.
- (1b) The mere simultaneous presence of *I*-final is what blocked *D*-final from causing *E*.

It is hard to see what other alternatives would make (1) true. But here's the rub: if (1a) is true, then (2) is false; and if, on other hand, (1b) is true, then (3) is false.

Suppose (1) is true in way (1a), that is, what blocks *D*-final's causing *E* is that *I*-final causes *E*. Then *D*-final's causing *E* is not blocked (it is not determined that *D*-final will not cause *E*) until there occurs at least the beginning of *I*-final's causing *E*. But the process of *I*-final's causing *E* does not begin until *E* begins: since *I*-final's causing *E* is indeterministic, nature does not determine that *I*-final does produce *E* until *E*'s being produced by *I* has begun to happen, and when *E*'s being produced has begun to happen so has *E* begun to happen. This means that, in the counterfactual scenario, where, because *I*-final fails to cause *E*, *D* does, it is not determined that *D*-final will cause *E* until the interval $[t1-t2]$ has begun without *E*'s having begun, that is, until after *t1*. In the counterfactual scenario *D*-final could not begin to cause *E* until after *t1*, and therefore the *E*-type event that *D*-final causes in the counterfactual scenario must begin slightly later than it does in the actual scenario, contrary to what (2) says.

If, on the other hand, (1) is true in way (1b) – if what does the preempting of *D*-final is, not *I*-final's causing *E*, but just *I*-final's occurring simultaneously with *D*-final – then (3) must be false; that is, *E*'s occurring at $[t1-t2]$ was avoidable by *S*. For to say that the mere simultaneous presence of *I*-final is what blocks *D*-final's efficacy is to say that *D*-final's being followed by *E* is nomologically necessary *only in the absence of simultaneous I-final*, that where *D*-final and *I*-final occur together, as in the actual scenario, they have the same real chance of not being followed by *E* as the chance that *I*-final will not indeterministically cause *E*. And that means that *E* at $[t1-t2]$ (*S*'s deciding at *t* to *A*) was no more unavoidable by *S* in the actual scenario than it would have been if the deterministic process had been absent. If it be insisted that *D*-final nomologically necessitates *E* whether or not *I*-final is present, then we have no reason to think that the presence of *I*-final preempts *D*-final, that *D*-final is not efficacious here. And a reason to think this is needed: we cannot just stipulate that it is so. If there is no such reason, then overdetermination of *E* by both *D*-final and *I*-final would seem the more plausible verdict (assuming that we have reason to think that *I*-final was also efficacious; if it's unclear what that reason

might be, then an at least equally plausible verdict in this case is that *D*-final preempts *I*-final).

The upshot is that, since Mele–Robb's example cannot satisfy all of (1), (2) and (3), it cannot succeed in escaping the dilemma that thwarts success in attempts to construct a Frankfurt-type counterexample to PAP: either the example fails to make it the case that *S* could not have avoided deciding at *t* to *A* or it makes that event's unavoidability depend on its being deterministically caused.

Notes

- Does Black's having set up a sufficient condition for Jones's doing *B* by *t3* mean that Black was responsible for the fact that Jones did *B* by *t3*? I'm inclined to think not. To my ear, saying that Black was responsible for that fact implies that, if it weren't for the sufficient condition that Black set up, that fact would not have obtained. And that is false. Jones also contributed a sufficient condition for it, his doing *B* at *t1*. These two sufficient conditions, Jones's and Black's, are independent of each other: either one could have occurred without the other and, had it done so, it would have ensured that Jones did *B* by *t3*. That means that neither Jones nor Black can be said to be responsible for that fact. But for each of them there is a morally equivalent fact for which he is responsible, namely, that he contributed a certain condition that was, in the circumstances, sufficient (logically or causally) for Jones's doing *B* by *t3*. In these respects this case is similar to some familiar puzzle cases that have been around for a long time in discussions of responsibility. Suppose that *A* and *B* simultaneously fire bullets through *C*'s head. We cannot say of either *A* or *B* that he is responsible for the fact that *C* died of a bullet wound to the head, because we can't say of either one that, but for his firing his bullet, *C* would not have died of a bullet wound to the head. But we can say of each that he is responsible for a morally equivalent fact, namely, a certain sufficient condition of *C*'s dying of a bullet wound to the head. Another example: suppose *C* goes on a long desert trek and, in order to ensure that he will die in the course of it, *A* drills a hole in *C*'s canteen and *B* puts poison in it. *C* dies as a result of drinking the poisoned water. *B* is responsible for that fact but we cannot say that *B* is responsible for *C*'s dying in the course of his desert trek because we can't say that if *B* hadn't poisoned the water then *C* would not have died on the trek. In that case *C* would still have died on the trek, but from lack of water. We can say of each of *A* and *B*, however, that he is responsible for doing something that, in the circumstances, would ensure that *C* died on the trek. In general, when two different agents have contributed independently sufficient conditions for the obtaining of a given fact, it seems that neither of them can be said to be responsible for *that* fact – since neither one's contribution was necessary for the obtaining of the fact – but each can instead be said to be responsible for contributing a sufficient condition for its obtaining.
- Peter van Inwagen (1978 and 1983) gets credit for being the first to suggest that PAP specified to action states of affairs can be defended against Frankfurt-style examples by distinguishing between more and less specific action states of affairs. Surprisingly, however, van Inwagen does *not* think that distinguishing between more and less specific states of affairs can serve to refute Frankfurt's argument against the thesis that one is morally responsible for acting in a certain way only if one could have acted otherwise. In fact, van Inwagen seems to think that Frankfurt's argument

against that thesis is successful. Van Inwagen distinguishes between one's action of a certain sort at a certain time and its being the case that one acted in that way at that time. And, for reasons that I do not understand, he thinks that one might be responsible for one's action while failing to be responsible for its being the case that one acted as one did. To me, however, it seems evident that one is responsible for one's action of a certain sort at a certain time if and only if one is responsible for its being the case that one acted in that way at that time. Van Inwagen might say that the obtaining of a state of affairs is always an entity distinct from, and more abstract than, any concrete particular event in virtue of which the state of affairs obtains, and that this ontological distinction opens up at least the logical possibility that an agent could be responsible for one and not the other. I am willing to grant that distinction, but then it seems to me that what we are interested in holding people morally responsible for is always the obtaining of states of affairs. Given that we've settled whether a person is responsible for the obtaining of a certain state of affairs, if there is a further question as to her responsibility for something that realized the obtaining of that state of affairs, then this question, if it is of any interest, is always about the obtaining of some more specific state of affairs.

- 3 The time t_0 – the precise time at which the occurrence or not of C determines whether or not Black's mechanism forces Jones to do B – must begin earlier than does the time t_1 . For Black's mechanism can finish detecting whether or not C occurs at t_0 (and begin causing Jones to do B if C does not occur then) only at some time after t_0 begins. So, if t_0 begins no earlier than t_1 , Black's mechanism could not make Jones's doing B begin to occur until some time after t_1 begins; and, therefore, the precise chunk of time that Black's mechanism would cause Jones's doing B to occupy would not be the same as the precise chunk that it actually occupies, t_1 . However, my argument below concerning C does not depend on this point; it goes through whether or not t_0 begins before t_1 .
- 4 This example is strongly reminiscent of John Locke's example (*Essay Concerning Human Understanding*, Bk II, ch. XXI, sec. 10) of the man who voluntarily stays in a room that, unbeknownst to him, he is locked into and therefore not at liberty to leave.
- 5 Alison McIntyre (1994) advocates a necessary condition for moral responsibility that is weaker than PAP but that still rules out examples (1)–(3) above as cases of moral responsibility, because it requires that the agent could have performed the action if she had decided to do so and this conditional does not hold in any of examples (1)–(3). However, her necessary condition does not require that the agent could have *decided* to perform the action and therefore it does not rule out another sort of case where the agent could not have performed the action. Suppose example (3) above altered as follows: there is a device attached to Smith's chair that will be activated should Smith either begin to try to stand up or so much as consider the possibility of standing up and will in that case cause Smith to reaffirm and maintain her intention to remain seated. (I go along with McIntyre's assumption that deciding to do a thing requires that one have antecedently considered the possibility of doing it.) Call this example (4). Here the device makes it the case that Smith is unable, not only to stand up, but even to decide to stand up. The nearest possible world where Smith does decide to stand up is one where the device is not present and there is nothing to prevent Smith from acting on such a decision. Thus McIntyre's necessary condition – that Smith could stand up if she decided to do so – is satisfied and her account of moral responsibility commits her to saying that in this example Smith is responsible for not standing up but in example (3) above Smith is *not* responsible for not standing up. (McIntyre does not explicitly consider any example like (3), but she

does say about an example essentially like (4) that the agent could have done the action if she had decided to and therefore is responsible for not doing it.) But I cannot see how the difference between examples (3) and (4) justifies this change in judgment. If example (3) is one where Smith cannot be held responsible for not standing up – as surely it is – then so must be example (4), and the reason in both cases is that the action of standing up was not open to Smith.

- 6 In a later paper (Fischer, 1987) and a book (Fischer, 1994) Fischer refines his account of responsibility for actions, saying that it requires that the 'mechanism,' the causal process, that produced the action was 'responsive to reasons' in an appropriate way. In a still later paper (Fischer, unpublished) he gives a new account of responsibility for *non*-actions, saying that this requires that the non-action issue from an appropriately reasons-responsive mechanism. This new account does not entail the asymmetry between actions and non-actions that was a feature of the earlier account: now alternative possibilities are never required for responsibility, for either actions or non-actions. There's no need to examine the details of this new account in order to give the following reason for thinking that it (or any other account that makes it sufficient for responsibility that an action [or non-action] is produced by a specified sort of causal process) is no more successful than the earlier account. It has a similar counterintuitive consequence, namely, that Jones will be responsible for its being the case that he does B by t_3 in the case where Black's mechanism *intervenes* to cause this, *provided* it does so by manipulating Jones's brain in such a way as to bring about the action via a process that is appropriately reasons-responsive (or that is of the specified sort).
- 7 For another defense of PAP against Frankfurt's argument see Widerker (1995). Widerker's defense is rather different from mine, but there are affinities between them at important points and I have been very much aided in my thinking about the matter by reading an earlier version of his paper and discussing it with him.
I wish to thank John Fischer for helping me to make my report of his views in Fischer (1986) more accurate. I have presented earlier versions of the present chapter at the University of Rochester, Monash University, La Trobe University, Victoria University, the University of Canterbury, the University of Hull, Bar-Ilan University, and the University of Tel-Aviv. I am grateful to the discussion-participants on those occasions for many helpful comments.

References

- Fischer, John (1986), 'Responsibility and Failure,' *Proceedings of the Aristotelian Society* 86, pp. 251–70.
Fischer, John (1987), 'Responsiveness and Moral Responsibility,' in F. Schoeman (ed.), *Responsibility, Character, and the Emotions*, Cambridge: Cambridge University Press.
Fischer, John (1994), *The Metaphysics of Free Will*, Aristotelian Society Series Vol. 14, Oxford: Blackwell.
Fischer, John (unpublished), 'Responsibility, Control, and Omissions'.
Frankfurt, Harry (1969), 'Alternate Possibilities and Moral Responsibility,' *The Journal of Philosophy* 66, pp. 828–39.
McIntyre, Alison (1994), 'Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency,' *Philosophical Review* 103, pp. 453–88.

- van Inwagen (1978), 'Ability and Responsibility,' *Philosophical Review* 87, pp. 201-24.
van Inwagen (1983), *An Essay on Free Will*, Oxford: Clarendon Press.
Widerker, David (1995), 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities,' *Philosophical Review* 104, pp. 247-61.