

Moral Responsibility and Buffered Alternatives

DAVID P. HUNT

According to the so-called Principle of Alternate Possibilities, or *PAP*, moral responsibility depends on there being alternate courses of action available to the agent. Not just any alternative will do (Pam's power yesterday to order steak instead of pasta is irrelevant to her responsibility today for tripping the Dean at the faculty meeting), though it need not be an alternative to the action for which responsibility is being attributed (Pam's responsibility for tripping the Dean won't rest on her power to refrain from tripping him if the deliberate extension of her leg across his path preempted a muscle spasm that would have produced the same result a split-second later, but it might still rest on her having been able to refrain from *intending* to trip him). This suggests the following formulation of the principle:

PAP: An agent X is morally responsible for performing an action A only if X did something B (which might or might not be identical to A) such that (i) X could have refrained from doing B and (ii) it is at least in part in virtue of X's having been able to refrain from doing B that X is morally responsible for performing A.

While different analyses of the relevant sense of "could" have led to very different ways of interpreting PAP, the principle itself (however interpreted) has inspired near unanimity across the philosophical spectrum. In a famous article, however, Harry Frankfurt argued that the principle is subject to counterexample.¹

1. Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66 (December 4, 1969), 829–39.

FRANKFURT COUNTEREXAMPLES

A “Frankfurt counterexample” to PAP has the following form. First, one is asked to consider a paradigm case of moral responsibility, one in which the putative agent and enviroing situation satisfy all one’s favorite criteria, whatever they might be (including the availability to the agent of alternate courses of action, if that is among the criteria). Second, a mechanism is introduced which strips the agent of all PAP-relevant alternatives while leaving everything else that makes the case a paradigm of moral responsibility exactly as it was. Finally, one is invited to see that the agent is still morally responsible, and that the availability or unavailability of alternatives, just by itself, is therefore irrelevant.

The viability of such a counterexample depends crucially on the nature of the mechanism, which must be both passive (allowing the agent to proceed unmolested down the path he is actually taking) and aggressive (vigilantly guarding against any significant deviation from this path). To bring off this delicate balancing act, the typical Frankfurt counterexample posits for its mechanism a “counterfactual intervener”: a device capable of interfering with an agent’s mental processes to ensure his acting in a certain way and primed to exercise this capacity if certain triggering conditions obtain, but which in the absence of those conditions simply “idles.” Consider a scenario *S* involving an agent *X* and an action *A* such that all of the following are true:

- (1) *X* is morally responsible for doing *A*;
- (2) *X*’s doing *A* satisfies the requirements for moral responsibility spelled out in PAP;

and

- (3) any action which plays the role of *B* in PAP—i.e., which is such that (i) *X* could have refrained from doing it and (ii) it is at least in part in virtue of *X*’s having been able to refrain from doing it that *X* is morally responsible for performing *A*—is such that *X* would refrain from doing it just in case some prior condition(s) *C* were to obtain.

Now *add* to *S* a mechanism *M* designed to detect *C* and to force *X* to do *B* if (and only if) *C* obtains. This recipe yields the classic Frankfurt counterexample to PAP.

To see why PAP might be thought threatened by such examples, consider what *M*’s insertion into the scenario implies for the original *S*. It would appear, in the first place, that (2) can no longer be true: given *M*, *X* has no alternative to doing *B*, since the only conditions under which *X* would *not* do *B* are conditions under which *M* brings it about that he *does* do *B*. (Since *M*’s addition to the scenario disrupts the specified connection between *C* and *B*, (3) will now be false as well.) But the truth of (1) is arguably *unaffected* by the introduction of *M*. The scenario into which *M* is being inserted is one in which *C* does not obtain (if it did obtain, *X* would not do *B*, and *ex hypothesi* *X* does do *B*). Because *C* does not in fact obtain, the necessary (and sufficient) trigger for *M*’s intervention is missing; *M* therefore enters *S* only as an idle spectator of the action. But then it’s hard to see why *M*’s introduction into the scenario should make any difference to (1). Con-

sider *Assassin*, the *Ur*-Frankfurt counterexample, in which Jones murders Smith on his own, but the existence of a prior sign of his decision to murder or not to murder allows Black to program a device to force Jones's decision to kill Smith if the sign goes one way and leave Jones alone if (as actually happens) it goes the other way. Frankfurt thought it obvious that Black's device, while successfully stripping Jones of alternative courses of action, left Jones's moral responsibility intact, for it "played no role at all in leading him to act as he did," "could have been subtracted from the situation without affecting what happened or why it happened in any way," is "irrelevant to the problem of accounting for [Jones's] action," and "does not help in any way to understand either what made him act as he did or what, in other circumstances, he might have done."² But if (1) is true and (2) is false, as appears to be the case in *Assassin*, PAP must be rejected.

Let us agree that adding M to S makes no difference to X's moral responsibility: if X is responsible for A in the original S, from which M is absent, then X is responsible for A in the new S, which differs from the original S only in that M is present (and whatever is implied by its presence). Frankfurt's reasons for this conclusion have considerable force, as even his critics concede. What is eminently disputable, however—and what published critiques of Frankfurt have in fact tended to dispute—is whether S is such that X can be responsible for A in the first place.

That there is a necessary and sufficient prior condition for X's doing otherwise, as stipulated in (3), is not an inconsequential feature of S. If C serves merely as a *prior sign* that X will refrain from B—generally reliable, perhaps, but neither necessary nor sufficient for the signaled outcome—M might fail to operate as advertised: for if, when C obtains, it is nevertheless possible that X might do B on his own, then M's introduction into S and consequent intervention to bring about B (automatically triggered by C) may be unnecessary; and (worse) if it remains possible, when C does *not* obtain, that X might nevertheless refrain from B, then M simply cannot deliver on its promise to eliminate all morally relevant alternatives. But the necessity and sufficiency of C, arguably essential to Frankfurt's counterexample, is at the same time an extraordinary proviso whose consistency with the rest of S is questionable. This prior condition either lies within X's power or it does not. The problem with S can therefore be presented in the form of a dilemma.

If it is *not* within X's power whether or not C obtains, it's hard to see how X could refrain from doing B in the original S, as required by PAP; for X's doing B is entailed by something (the non-obtaining of C) over which X has no control. In that case (2) and (3) are incompatible with each other, and S does not constitute a possible scenario. At least it does not constitute a possible scenario without recourse to a conditional analysis of X's power to do otherwise, rejection of the relevant "power-entailment principle," or some other compatibilist stratagem. But Frankfurt's argument specifically disallows any such move.³ The principle Frank-

2. Frankfurt (1969, 836–7).

3. It *especially* disallows any reliance on a compatibilist interpretation of PAP. Since Frankfurt ends his essay by reviewing some reasons why his critique of PAP favors compatibilism, it would be unfortunate if his critique presupposed compatibilism all along.

furt undertakes to challenge in his essay is neutral on the question “whether someone who accepts it is thereby committed to believing that moral responsibility and determinism are incompatible”; it is instead a generic version of PAP, which “[p]ractically no one . . . seems inclined to deny or even to question . . . People whose accounts of free will or of moral responsibility are radically at odds evidently find in it a firm and convenient common ground upon which they can profitably take their opposing stands.”⁴ Clearly a critique of PAP that would succeed against compatibilists while begging the question against incompatibilists is not the argument Frankfurt intended to offer. Of course one possible outcome of the dispute over Frankfurt-type cases is that Frankfurt was wrong about this: *compatibilists* must reject PAP but *incompatibilists* can consistently retain it. (Incompatibilists could be forgiven for pointing out that this is just what they’ve been saying all along!) But if Frankfurt’s original argument is to succeed *as intended*, without begging the question against incompatibilists, S cannot be a scenario in which C lies outside X’s power, as incompatibilists understand “power.”

So much for the first horn of the dilemma. If, on the other hand, it *is* within X’s power whether or not C obtains, then the original S (prior to the introduction of M) includes an alternative: either C itself (if C is one of X’s possible actions) or some action C* which, were X to perform it, would bring it about that C obtains. If this alternative is morally significant, (3) can be compatible with (2) after all, and the coherence of S has been salvaged. But then M cannot do its job. The new alternatives, posited as a condition of S’s coherence, cannot be eliminated by M, since M is programmed to intervene only *after* C obtains. No intervener is quick enough to cancel the very conditions that triggered its intervention.

The problem, in sum, is that if it *is* within X’s power whether or not C obtains, then S contains an alternative that M cannot preclude; and if it is *not* within X’s power whether or not C obtains, X does not appear to be morally responsible in the first place. In neither case does S + M constitute a scenario such that an open-minded opponent could be expected to agree that X is morally responsible yet lacks all morally relevant alternatives; so S + M fails to provide a counterexample to PAP.⁵

Defenders of Frankfurt’s argument have offered various replies to this dilemma, most of which allow that the presence or absence of the conditions triggering the mechanism constitutes a residual alternate possibility, while insisting that scenarios are available in which this possibility is morally irrelevant. John Fischer, for example, appeals to cases in which X would exhibit some involuntary signal, like a blush or a twitch, were he about to decide against A-ing, while Eleonore Stump develops a counterexample in which a decision is physically realized by a complex pattern of neuron firings whose initial segment is sufficient to identify the incipient mental event as a *decision to do* Φ though the entire sequence

4. Frankfurt (1969, 829).

5. David Widerker is particularly notable for deploying this sort of dilemma against the Frankfurtian—for example, in his “Libertarianism and Frankfurt’s Attack on the Principle of Alternative Possibilities,” *Philosophical Review* 104 (April 1995), 247–61. See also Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), 142–43.

must be completed for the decision to occur, and the mechanism is set to intervene if (and only if) it detects the opening pattern signaling the “wrong” decision.⁶ If the critics object that such cases involve alternatives that M cannot preclude—the possibility that X might have blushed (though he didn’t), or that certain neurons might have fired (though they didn’t)—the Frankfurtian can dismiss these alternatives as mere “flickers of freedom” which fail to satisfy clause (ii) of PAP. Fischer has even suggested that the “moral irrelevance” strategy can be deployed on behalf of the dilemma’s other horn, which alleges that S presupposes causal determinism: if $C/\sim C$ is *not* an alternate possibility for X, it is nevertheless too exiguous, and therefore too inconsequential in our intuitive assessment of the case, to *undermine* X’s moral responsibility under PAP, just as it is too exiguous to *support* X’s moral responsibility if $C/\sim C$ is an alternate possibility.⁷

I do not propose to evaluate these Frankfurtian rejoinders here. I believe that they have some force, and may well establish that it is *not irrational* to reject PAP on the basis of Frankfurt’s argument. The problem is that these elaborations on Frankfurt’s original counterexample are unlikely to compel assent from anyone not already persuaded, especially those with strongly incompatibilist intuitions.⁸ While partisans on both sides may be reluctant to declare a stalemate, it does not appear that further belaboring the classic Frankfurt counterexamples will yield much progress at this stage in the debate. It is therefore worth considering whether an alternative to the classic counterexamples might be more effective.

BLOCKAGE AND BUFFERS

The dilemma impeding the successful prosecution of Frankfurt’s original argument appears to turn on the status of the conditions triggering the mechanism’s intervention. The fact that the alternative-eliminator in the classic Frankfurt counterexamples interferes *only* when triggered plays a crucial role in persuading us that the agent is morally responsible in the actual scenario, in which the triggering conditions do not obtain and the mechanism simply idles; but it’s this very dependence on a trigger that prevents the mechanism from eliminating *all* alternatives, at least in scenarios which do not simply beg the question by presupposing a deterministic environment.

The most radical response to this difficulty would be to devise a counterexample to PAP that dispenses with triggering conditions altogether. Elsewhere I

6. For Fischer, see his *The Metaphysics of Free Will* (Oxford: Blackwell Publishers, 1994), chap. 7; for Stump, see her “Libertarian Freedom and the Principle of Alternative Possibilities,” in *Faith, Freedom and Rationality*, ed. Daniel Howard-Snyder and Jeff Jordan (Lanham, MD: Rowman and Littlefield, 1996), 73–88.

7. John Martin Fischer, “Recent Work on Moral Responsibility,” *Ethics* 110 (1999), 113.

8. For a similar assessment of the dialectical situation, see Derk Pereboom, “Source Incompatibilism and Alternative Possibilities,” in *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, ed. David Widerker and Michael McKenna (Aldershot, UK: Ashgate Pub. Ltd., 2003), 190–93.

have proposed a so-called “blockage” case that attempts to do just this.⁹ The basic idea is inspired by John Locke’s famous example in which “a man be carried whilst fast asleep into a room where is a person he longs to see and speak with, and be there locked fast in, beyond his power to get out; he awakes and is glad to find himself in so desirable company, which he stays willingly in, i.e. prefers his stay to going away.”¹⁰ Here the alternative-eliminating locked door does not become locked when certain triggering conditions are satisfied and remain unlocked otherwise; it is simply locked. The agent’s freedom in the example does not depend on the door’s being unlocked in the actual sequence of events, but on the fact that the agent never (freely) comes up against the locked door. In my example the walls of the room are brought right inside the man’s mind: Jones, in the blockage version of *Assassin*, decides on his own, but through a tremendous coincidence Black’s fiendish mechanism has actually (not counterfactually) “locked” all the neural pathways other than the ones Jones is actually using. Since Jones decides on his own, he is arguably morally responsible.¹¹

Unfortunately, while blockage cases have much to recommend them, they come with their own set of difficulties, and recent discussion of my proposal by John Fischer, Derk Pereboom, Robert Kane, and David Widerker (among others) confirms my own reticence about pushing them too strongly as definitive counterexamples to PAP, at least in their present form.¹² The central difficulty is that

9. I first presented this alternative to the standard Frankfurt counterexample in “Moral Responsibility and Unavoidable Action,” *Philosophical Studies* 97 (January (II) 2000), 195–227. In “Freedom, Foreknowledge and Frankfurt,” in Widerker and McKenna, pp. 159–83, I develop such counterexamples in greater detail, referring to the non-counterfactual mechanisms at work in them as “passive alternative-eliminators.” John Fischer, however, has dubbed them “blockage cases,” and the name has stuck.

10. *An Essay Concerning Human Understanding*, Bk. II, chap. XXI, section 10.

11. Others who endorse this strategy include Michael Zimmerman, whose “The Moral Significance of Alternate Possibilities” features a counterexample modeled closely on my own, and Michael McKenna, who develops and defends a variation on my example (he calls it “limited blockage”) in “Robustness, Control, and the Demand for Morally Significant Alternatives: Frankfurt Examples with Oodles and Oodles of Alternatives.” Both these papers may be found in Widerker and McKenna, pp. 301–25 and pp. 201–17, respectively. John Fischer has also given qualified support to the blockage approach, while astutely identifying some of its vulnerabilities in his “Recent Work on Moral Responsibility.” Another counterexample that is often treated as a species of blockage, though it is significantly different from the kind of example I offer, is Al Mele and David Robb’s much-discussed “preemption” case, first presented in “Rescuing Frankfurt-Style Cases,” *Philosophical Review* 107 (1998), 97–112. Robert Kane distinguishes the two species of blockage by calling the Hunt cases “pure blockage” and the Mele/Robb cases “modified blockage”; see his “Responsibility, Indeterminism and Frankfurt-Style Cases: A Reply to Mele and Robb,” in Widerker and McKenna, pp. 91–105. Yet another kind of counterexample associated with blockage is divine foreknowledge—or so I argue in “Freedom, Foreknowledge, and Frankfurt,” also in Widerker and McKenna, pp. 159–83.

12. John Martin Fischer, “Recent Work on Moral Responsibility”; Derk Pereboom, “Alternative Possibilities and Causal Histories,” *Philosophical Perspectives* 14: Action and Freedom (2000), 119–37; Robert Kane, “Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson,” *Philosophy and Phenomenological Research* 60 (2000), 157–67; David Widerker, “Theological Fatalism and Frankfurt Counterexamples to the Principle of Alternative Possibilities,” *Faith and Philosophy* 17 (April 2000), 249–54.

the conditions barring Jones's access to alternative pathways and guaranteeing his decision to kill Smith must be distinguished, in some non-*ad hoc* way, from the sorts of conditions that would beg the question against incompatibilism by *causally determining* Jones's decision.¹³ While blockage cases do seem intuitively different at some level from cases of straightforward causal determination, it's not clear that defending the intuition against convinced PAPists will be any easier than tamping down "flickers of freedom" in the classic Frankfurt counterexamples.¹⁴

So blockage is probably not the magic bullet for which PAP's critics are looking. There is, however, a less radical alternative to the conventional Frankfurt counterexample, one that reconceives the triggering mechanism rather than abolishing it. Counterfactual triggers have typically been thought of as conditions *sufficient* for the agent to perform the alternative action (absent intervention by the mechanism). The triggers in *Assassin*, for example, are of this sort. But the mechanism is just as effective at excluding the requisite alternatives when it is triggered by *necessary* conditions. Here is an example—call it *Revenge*.

Jones, while attending a party, is deliberately humiliated by Smith. The first thought that occurs to Jones, after realizing what Smith has done to him, is that he would like to kill Smith. He leaves the party, escaping the immediate pressures of the situation and giving himself ample opportunity to pull back from this line of thinking. Given the kind of person Jones is, and given the situation in which he finds himself, the alternative of *not* killing Smith is not unthinkable for him; moreover, should he entertain this alternative, nothing would prevent him from deciding and acting on it. But Jones could decide (and act) otherwise only if he first considered acting otherwise, and he never does this (though he could); instead, he nurses his grievance without respite, while the idea of killing Smith becomes more

13. I raise this worry, and offer a tentative defense against it, in "Freedom, Foreknowledge, and Frankfurt." All the sources cited in note 12 contain some critical discussion of it, in one form or another. Note that divine foreknowledge, employed as a blockage counterexample to PAP, escapes this objection, since it's divine *infallibility* rather than *causal determination* that eliminates alternatives. Unfortunately, this advantage for the divine foreknowledge version of blockage is undermined by the possibility that the foreknowledge in question is incoherent.

14. Al Mele and David Robb offer a vigorous defense against this charge in "Bbs, Magnets and Seesaws: The Metaphysics of Frankfurt-Style Cases," in Widerker and McKenna, pp. 127–38. As a fellow "blocker," my intuitions are very much like theirs on this subject. I nevertheless find some justice in Derk Pereboom's concluding comments on my version of blockage:

Sympathy for Frankfurt-style arguments is generated by the sense that moral responsibility is very much a function of the features of the actual causal history of an action, to which restrictions that exist but would seem to play no actual causal role are irrelevant. However, in a scenario in which such restrictions, despite initial appearances, could be relevant to the nature of the actual causal history of an action after all, one's intuitions about whether the agent is morally responsible might become unstable. My own view is not that actual causal histories in blockage cases are clearly deterministic, but only that these considerations show that they may be. This type of problem should make one less confident when evaluating these difficult kinds of Frankfurt-style cases. Since Mele and Robb's development of their case involves something very much like a blockage scenario, one might as a result also be less confident about the ultimate success of their argument.

and more attractive to him. Having finally decided to do the deed, he gets a gun from his car, returns to the party, and shoots Smith dead.

Before introducing the counterfactual intervener, let us note some of the salient features of *Revenge* as it has been described up to this point. (Of course any and all of these features may be contested later—this is just a provisional inventory.) In the first place, the idea of killing Smith comes to Jones unbidden (so the question whether he can do otherwise in *this* respect—that is, *not* have the idea instead of having it—does not even arise).¹⁵ In the second place, once the idea has come before his mind, it is open to Jones to embrace the idea, thereby deciding to kill Smith, and it is open to him to dismiss the idea, thereby deciding not to kill Smith (so nothing constrains him one way or the other). Thirdly, given Jones's psychology and given the circumstances in which he finds himself (including the fact that he is already thinking about killing Smith), he can avoid deciding to kill Smith only if he first considers the alternative of not killing Smith (so his considering not killing Smith—minimally, the conscious awareness, however brief, of *not killing Smith* as a possibility for him—is a *necessary condition* for his accessing an alternative to his actual decision). Finally, Jones decides to kill Smith without ever considering the alternative of not killing Smith (so he makes his decision without deliberating). There is, I would suggest, no *prima facie* reason for regarding these features of *Revenge*, either jointly or severally, as incompatible with Jones's moral responsibility. Of course *secunda facie* concerns might arise on reflection, and some of these features will in fact figure in objections to be considered later. But pending the successful development of such objections, we are surely within our rights in supposing *Revenge*, at least as it has been characterized up to this point (prior to the introduction of the mechanism), to be a case in which Jones is morally responsible for the murder of Smith.

The final element to be added to *Revenge* is the counterfactual intervener, which differs from the device in *Assassin* inasmuch as it is programmed to hijack Jones's mental processes and force him to decide to kill Smith if he so much as *considers* not killing Smith. With this device in place, there is no alternative to Jones's deciding to kill Smith: Jones can decide otherwise only if he first considers doing so, but then the device will force him to decide *to* kill Smith. So an alternative decision is not available to Jones in *Revenge*; and if this prime candidate for a PAP-relevant alternative can be excluded by the mechanism, it would appear

15. It might be objected that Jones could have done something about being *the kind of person* to whom such a thought would come unbidden in the specified circumstances; it's therefore not true that he could not have avoided having this thought, for he *would* have avoided it if he had taken advantage of alternatives that were available to him in earlier character-forming episodes in his life. This objection can be circumvented with a slight modification to *Revenge*. Suppose the thought of killing Smith is planted in Jones's mind by an external source—perhaps even by Black's mechanism. So its coming to be there has nothing to do with Jones being one kind of person rather than another. But given that it's there, Jones's duty is now to accept or reject this idea, and nothing constrains him one way or the other. (A related objection appeals to earlier choices that formed Jones into the kind of person who would *accept* the idea of killing Jones, once the idea occurred to him. This objection cannot be handled by a simple revision in the terms of *Revenge*. For my response, see the third objection and reply in the section, "Three Objections Claiming the Presence of Alternatives.")

that any plausible candidate can be similarly excluded. Yet Jones, who in fact proceeds to murder Smith on his own, leaving the device untriggered, seems morally responsible for killing Smith. PAP is therefore wrong when it makes moral responsibility rest on the availability of morally relevant alternatives.¹⁶

Revenge differs from *Assassin* in that there is a “buffer-zone” between the agent’s actual pathway and any alternative pathways which might plausibly satisfy PAP. Entering the buffer-zone triggers the intervener, so the PAP-relevant alternatives are out of reach. In such “buffer cases” the relevant alternatives are all “buffered alternatives.” *Revenge* is a buffer case, and buffer cases appear to escape the usual objections to Frankfurt counterexamples like *Assassin*, which rest on the supposition that the prior signs to which the mechanism is responsive constitute *sufficient* conditions for the agent’s performing or refraining from B (absent intervention by the mechanism).

In the remainder of this paper I defend the “buffer strategy” by considering and responding to six counterarguments. The basic worry underlying these objections is that, appearances notwithstanding, either non-buffered alternatives *are* available in *Revenge*, or Jones is not really morally responsible.

THREE OBJECTIONS CLAIMING THE ABSENCE OF MORAL RESPONSIBILITY

The first set of objections presupposes that *Revenge* successfully eliminates all of Jones’s morally relevant alternatives; at least it concedes, if only for the sake of argument, that *Revenge* can be so interpreted. What it denies, however, is that Jones is then morally responsible for deciding to kill Smith. Insofar as *Revenge* is a case in which Jones genuinely lacks the requisite alternatives, it is a case in which he is not morally responsible.

Objection 1. If Jones’s considering the possibility of not killing Smith is a causally *necessary* condition for his refraining from deciding to kill Smith, then his *not* considering this possibility is a causally *sufficient* condition for his deciding to kill Smith. In *Revenge*, where Jones does not in fact consider the possibility of desisting, this condition appears to be satisfied. The scenario in *Revenge* therefore rules out the causal indeterminism that libertarians make a prerequisite for moral responsibility. Buffer cases, it would appear, cannot so easily escape the question-begging sufficient conditions which jeopardize the effectiveness of the original Frankfurt counterexamples.¹⁷

Reply. This objection fails on two counts. In the first place, causally sufficient conditions alone do not undermine moral responsibility, even for a libertarian, so long as it’s the *agent* who determines whether those conditions obtain. Jones could not, for example, disclaim responsibility for the murder by noting that the force

16. This is an elaboration of a case I first presented in “Moral Responsibility and Unavoidable Action,” where it took a backseat to my development of the blockage strategy. Derk Pereboom has independently adopted a similar line: his *Tax Evasion*, like *Revenge*, employs necessary rather than sufficient conditions as the mechanism’s trigger. See Pereboom’s *Living without Free Will*, pp. 18–37.

17. David Widerker has suggested this objection to me in correspondence.

and trajectory of the bullet, once the gun had been fired, were causally sufficient for Smith's death, since it was up to Jones whether to pull the trigger. Likewise Jones cannot escape responsibility on the grounds that his not considering an alternative course of action was causally sufficient for his deciding to kill Smith, since it was up to Jones whether to consider an alternative.

In the second place, however, the objector has simply failed to identify a causally sufficient condition in the example, let alone one that might undermine moral responsibility. What *Revenge* stipulates as a necessary antecedent condition for Jones's refraining from deciding to kill Smith is that Jones *at some time or other* consider the alternative of not killing Smith. The sufficient condition for his deciding to kill Smith which corresponds to this necessary condition for refraining is not *there being some time or other* at which he *does not* consider not killing Smith, but *there being no time at all* at which he *does* consider not killing Smith. This condition, however, is never available as a causally determining factor in *Revenge*, since at any time prior to deciding to kill Smith, Jones can always falsify the condition by considering not killing Smith.

Objection 2. Because Jones never considers any alternative(s) to killing Smith, he can't really be said to have *chosen* to murder Smith. Consequently *Revenge* is not a paradigm case of moral agency. Any test for moral responsibility that *Revenge* passes with flying colors should be greeted with suspicion, and it is a virtue of PAP, not a vice, that it withholds its imprimatur in this case.¹⁸

Reply. This objection would garner little support in any other context. To begin with, it's far from clear that paradigmatic moral responsibility *does* require "genuine" choice, by which the objector appears to mean a choice in which at least one of the unchosen alternatives was actively entertained prior to the choice. Why think that choice, in this sense, is a requirement? Perhaps the idea is this: considering an alternative assures the agent that he *has* alternatives, and without such assurance the exercise of agency is pointless. Or maybe the idea is this: moral agency is *intentional*, undertaken with *knowledge* of what one is doing, and such knowledge, to play its action-guiding role, must be sufficient for distinguishing one's intended action from other things that one does not intend to do. The key idea in either case is that moral agency, in the full and paradigmatic sense, must involve *contrastive* knowledge of one's actions.¹⁹ But this is dubious: surely, I can presume that I have alternatives in general without considering any one of them in particular, and I can know what I am doing without exercising the capacity to know what I am *not* doing. So there is no good reason to accept the suggested requirement; and even if it were accepted, there is no good reason to think the requirement couldn't be satisfied by a *tacit* or *dispositional* knowledge of alternatives, as opposed to the full-blown, conscious canvassing of options that the objec-

18. The mirror image of this objection is that, since Jones clearly *is* morally responsible in *Revenge*, there *must be* a genuine choice or set of choices that accounts for Jones's responsibility; and since such choice has been excluded from the immediate scenario, it must lie somewhere in Jones's past. See the third objection in the section "Three Objections Claiming the Presence of Alternatives."

19. Michael McKenna raised this general worry in his response to an early version of the paper, presented at the Eastern Division meeting of the American Philosophical Association in 2000.

tor takes to be so important. Indeed, some choices may be such that seriously considering an alternative would reveal a moral defect, and the moral credit due the agent for finally making the right decision would be correspondingly *diminished*.

But even if the central paradigm of moral responsibility *does* feature an explicit awareness of alternatives, PAP's business is to lay down a necessary condition, not to elaborate a paradigm. If Jones, on the way to the party, had instinctively pushed a child out of the path of an oncoming vehicle, without first entertaining the alternative of not helping the child, he would still be morally praiseworthy for his action. Likewise he is morally blameworthy for killing Smith even though he never considered not killing him. Perhaps *free will* requires deliberation between alternatives; perhaps, as Peter van Inwagen has argued, it even requires deliberation of a very rare and special sort.²⁰ But whatever the merits of this account of free will, there is no reason to think it should be extended to moral responsibility as well.

Objection 3. If Jones's moral responsibility is not undermined by something he *doesn't* do (engage in genuine deliberation over whether to kill Smith), it is undermined by something he *can't* do (immediately access the alternative to his actual decision). The imposition of this necessary condition on his deciding otherwise involves a serious—perhaps even an unacceptable—limitation on his free agency. If he really did labor under this limitation, as *Revenge* stipulates, we should expect his moral responsibility to be similarly limited. But then we shouldn't be surprised that the case does not satisfy PAP.

Reply. That one step in practical reasoning is sometimes necessary for another step to be taken is part of what makes our mental lives intelligible; the fact that our minds exhibit such structure should not count against free agency or moral responsibility. So the stipulation in *Revenge* that Jones must consider refraining before he can choose to refrain can hardly be dismissed as an arbitrary and adventitious limitation on Jones's powers. There is no need to resort to science fiction stories here.

Note that it is not necessary, for the example to succeed, that *everyone's* psychology be so structured. Perhaps we can even imagine an agent whose psychology provides no purchase at all for Black's device. Consider, for example, someone like Dostoevsky's "underground man"—obsessed with his own autonomy, hostile to the "tyranny" of reason, and allergic to any essentialist account of himself that would threaten to turn him into a predictable, clockwork mechanism; and imagine that this individual has succeeded in freeing his volitional life from subjection to rational norms and psychological habits, with the result that he is now completely *spontaneous*: able to form intentions immediately, without traversing any intervening steps, including a prior consideration of their propositional contents. All this would show, however—supposing such a bizarre specimen to be possible—is that a different agent than Jones and a different scenario than *Revenge* might be

20. Peter van Inwagen, "When Is the Will Free?" reprinted in *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. Timothy O'Connor (New York & Oxford: Oxford University Press, 1995), pp. 219–38. But he concludes, in agreement with the last sentence of this paragraph, that "the radically limited domain of the freedom of the will" does not imply "a similarly radically limited domain for moral blame" (p. 238).

hard to buffer. A putative counterexample to PAP is not disqualified just because other cases are not counterexamples. For *Revenge* to succeed, it is only required that *someone's* psychology (might) be structured in the way indicated, and that its being so structured does not limit the person's freedom or responsibility.

Even if such structuring is psychologically plausible and consistent with moral responsibility, however, it might still be questioned whether it is strong enough to exclude the very *possibility* that Jones might decide against murder without first considering this alternative; some, such as Sartre, might even regard this possibility as integral to human freedom. Suppose then, for the sake of argument, that Sartre is right about the "human condition": that I am choosing every moment what I will be; that I can therefore choose this instant (and so without prior deliberation) to become someone quite different—say, to abandon my family and academic career and become a circus roustabout; and further, that the availability of this alternative possibility helps explain why I am morally responsible for persisting in my more conventional lifestyle. Nevertheless, it just does not follow (or so it seems to me) that if I *could not* access this alternative except *via* some intermediate step (which might range from merely considering it to engaging in a full-scale review of my values and priorities), I would not be morally responsible for my present choices. To insist on the perpetual possibility of 180-degree turns in one's moral life is an unreasonable requirement for both free agency and moral responsibility.

As a final thought (giving in now to the lure of science fiction), imagine that astronauts exploring an alien planet encounter an intelligent life-form and wonder whether to regard it as a moral agent. There are things they might find out about this life-form that would lead them to withhold this attribution, but discovering that it couldn't decide to A unless it first considered A-ing would not be among them. Or suppose neurophysiologists investigating the mechanism by which *we* form intentions find out, as a matter of brute empirical fact, that our own mental "wiring" permits the acquisition of an intention only after the intentional object has come before the mind. Would this scientific discovery have any tendency to overturn our conception of ourselves as moral agents? I doubt that it would. (There surely *are* conceivable scientific discoveries that would cast doubt on human freedom. But this is not one of them.) Whether or not Sartre is in fact right about the human condition as we find it, "Sartrean freedom" is not a prerequisite for moral responsibility, and supposing that Jones lacks it does not disqualify him from moral agency.

THREE OBJECTIONS CLAIMING THE PRESENCE OF ALTERNATIVES

The next set of objections presupposes that Jones is, or can be construed to be, morally responsible in *Revenge*; but it insists that morally relevant alternatives, of one sort or another, will then be present. Insofar as *Revenge* is a case in which Jones is genuinely responsible for his actions, it is a case in which alternatives sufficient to satisfy PAP are in fact available.

Objection 1. In *Revenge*, as in *Assassin*, the counterfactual intervener cannot act quickly enough to squelch the triggering event itself. So this event—in *Revenge*,

Jones's considering not killing Smith—remains an open possibility. But then Jones, who in actuality does *not* consider not killing Smith, has available to him an alternative course of action which the pro-PAP party can exploit. It was the failure of the traditional Frankfurt counterexamples to snuff out all “flickers of freedom” that led us to seek alternatives to *Assassin* and its ilk. If *Revenge* can't eliminate the flickers, its alleged advantage over *Assassin* disappears.

Reply. Whatever might be said of the residual alternatives in traditional Frankfurt counterexamples like *Assassin*, the alternative left open in *Revenge* does not satisfy clause (ii) of PAP. If, as in *Revenge*, an agent is morally responsible in virtue of being *blameworthy*, a PAP-relevant alternative should be such that, by accessing it, the agent would avoid (or mitigate) blame; but merely considering the possibility of not killing Smith would have no effect by itself on Jones's blameworthiness. This flicker fails to catch.

Compare this case with *Smoker*, in which Jack realizes that his first-order desire for cigarettes stands a good chance of yielding to his second-order desire to quit smoking if (but only if) he joins a 12-step program, while Mack, a smoking buddy who fears he will be left to smoke alone if Jack quits, stands ready, should Jack join the program, to slip him a pill which would make his craving for nicotine irresistible. Then Jack cannot access a decision to quit smoking, but he can (moved by the better angels of his second-order desire) decide to join the 12-step program. A defender of PAP might plausibly appeal to this alternative's availability to explain Jack's moral responsibility in the case where he doesn't join the program and continues to smoke “on his own,” without the ministrations of Mack's pill. But the appeal is plausible only on the assumption that Jack, in the alternative sequence of events, joins the program *as a means toward quitting*; this is what connects, “in the right way,” this possible alternative with the action for which responsibility is being attributed. (Contrast this with a case in which he joins the program in order to make the acquaintance of an attractive woman who attends the meetings: *this* alternative possibility, accessed for this reason, would not satisfy PAP.) In *Revenge*, however, undertaking to consider not killing Smith is something that Jones does not do *as a necessary means toward* the goal of not killing Smith. Indeed, it would appear that undertaking to consider not killing Smith *could not* play this role for him, inasmuch as it would impose on him the incoherent task of considering not killing Smith (in order to take not killing Smith as his intended end) *before* undertaking to consider not killing Smith.²¹ This candidate for a PAP-relevant alternative cannot get past clause (ii).

We might give this objection more credit than it's due because we think Jones *should* have considered not killing Smith: we blame him, not only for deciding to kill Smith, but also for not even considering not killing him. Since we hold him responsible for this failure, and an alternative was available to him (nothing

21. Or if not outright incoherent, it will involve two (possibly continuous) considerings, the second of which would be morally significant (because done as a means toward deciding not to kill Smith) but not available to Jones, since the first (PAP-irrelevant) considering would trigger the intervener.

prevented Jones from considering not killing Smith), *Revenge* appears to contain at least one action that satisfies PAP.²²

If this suggests that PAP remains alive and well, however, the suggestion is mistaken. In the first place, the fact that one morally appraisable constituent of *Revenge* (Jones's failing to consider sparing Smith's life) does not violate PAP does not show that another morally appraisable constituent (Jones's deciding to kill Smith) does not violate PAP, and a single counterexample is all that's needed. In the second place, PAP's failure in one case undermines its authority in other cases. If Jones is morally responsible both for deciding to kill Smith and for failing to consider not killing Smith, there should be some unitary account of what it is to be morally responsible that would encompass both cases. Since being-able-to-do-otherwise is unavailable for this role in the former case, it is doubtful that it is available in the latter; instead, whatever substitute for PAP turns out to provide the right account of Jones's responsibility for deciding to kill Smith (e.g., his being the uncaused cause of the decision, its issuing from a "reasons-responsive" mental process, and so on²³) will presumably provide the right account of his responsibility for the way in which he reached that decision, including his failure even to consider not killing Smith. Though the cleanest counterexample to PAP would contain no morally appraisable constituents for which there exist morally relevant alternatives, it may be impossible to devise such a counterexample without begging the question, and defeating PAP does not in any case depend on producing such a counterexample.²⁴

Objection 2. It is simply not true that Jones's only immediate options, given that he is currently thinking about killing Smith, are either bringing this thought to fruition by deciding to kill Smith, or else replacing this thought with the alternative thought of not killing Smith. Jones might, for example, simply *fail* to decide to kill Smith, without ever considering not killing him; in this case the triggering event would not occur, and Jones would successfully avoid killing Smith.

Reply. To count as an alternative satisfying PAP, Jones's refraining from deciding to kill Smith must be something that Jones *does*; it can't be something that is done to him, or that simply happens. It is the former kind of refraining for which Jones's considering not killing Smith serves in *Revenge* as a necessary condition. Now any case of Jones's not deciding to kill Smith will be a case of Jones's refraining (in this sense), or it won't be. If the latter—as when, for example, Jones fails to decide to kill Smith because he is himself killed, or is overcome by sleep, or is distracted by a loud explosion (after which he forgets what he had been thinking about)—the alternative is of the wrong sort to satisfy PAP. If Jones is morally responsible for killing Smith, it isn't in virtue of *these* alternative possibilities, the

22. Daniel von Wachter pressed this objection when I presented the paper at Oxford in 2003.

23. For an example of the first, see Eleonore Stump, "Alternative Possibilities and Moral Responsibility: The Flicker of Freedom," *Journal of Ethics* 3 (1999), 299–324; for an example of the second, see John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998).

24. Audiences at oral presentations of the paper have been more likely to raise this objection than any of the other five. The last paragraph of my reply, added late in the paper's development, is an attempt to understand why this objection is so resilient.

actualization of which would simply make Jones the beneficiary of “moral luck”; it is therefore unnecessary to set up a buffer zone around such alternatives when constructing a counterexample to PAP. But if the alternative is a genuine case of refraining, it’s not clear why we can’t simply stipulate in setting up the example that Jones can move to this alternative from his present course of action only by first considering this alternative. Whether or not any actual agents are so constituted, supposing that Jones is so constituted does not disqualify him as a moral agent. But then the mechanism can intervene to prevent Jones from accessing this alternative.

Perhaps the *does/done-to-him* distinction on which this initial reply rests is too artificial to cover all the alternatives the objector has in mind, making it possible to pass between the horns of the dilemma. Let us therefore consider one way things might go that does not fit neatly into this rubric. It would seem that Jones could leave the decide-to-kill-Smith track, not just *via* a (bufferable) consideration of the possibility of not killing Smith, or under the (PAP-irrelevant) prompting of an intervening cause, but in direct *reaction against* his contemplated course of action. Suppose, for example, he returns to the party, approaches Smith with gun drawn, and suddenly realizes what it is he is about to do; he then flees in horror. Here he moves to the alternative pathway, not because he is attracted to it (he couldn’t be attracted to it without first considering it, and this would trigger the intervener), but because he is repelled by what he is actually contemplating. There is some plausibility to the claim that he could *and should* have been affected in this way. The mechanism can certainly be reset to prevent Jones’s reaching a contrary *decision* along this new route, but it can’t intervene until Jones takes the first step along this route, and the fact that this first step is open to Jones seems morally relevant to his culpability for murder. At least it seems *more* relevant than the alternative of merely considering not killing Smith. For one thing, *feeling bad* about a prospective course of action represents a real if minimal departure from moral neutrality, whereas simply *entertaining an option* does not (by itself) commit the agent, even incipiently, to one side or the other. What’s more, if Jones *were* to go on to decide against murdering Smith, or to refrain from deciding to murder Smith, his discomfort at the thought of killing Smith might help explain why he didn’t decide to kill Smith, whereas his having considered not killing Smith could not similarly explain why he opted for this outcome rather than the other.²⁵

The problem with this reply is that we still need to understand how the availability of this overlooked alternative might satisfy PAP. It’s hard to see how it could do so unless the alternative in question—feeling bad, entertaining a doubt, making a negative judgment—*ipso facto* put any decision to kill Smith in abeyance. But it obviously does no such thing: any negative reaction to Smith’s murder on Jones’s part might be outweighed in Jones’s mind by other factors; or Jones’s trajectory toward action might be akratically impervious to moral qualms, prudential calculations, and the like. PAP’s defenders need an alternative that would allow the

25. The basic objection conveyed in this paragraph was suggested to me (independently) by Bob Adams and Gordon Pettit.

agent to avoid blame, and feeling bad or conflicted about what one is doing simply fails to do the job. So even assuming that Jones *could* react in the described way, acquiring a negative attitude toward his proposed action in one simple step, it's doubtful that this would rescue PAP. But one might also challenge the assumption. Let *Revenge* be fine-tuned so that Jones's responding in the suggested way would be completely out of character for him: someone who responded in this way just wouldn't be *Jones*. Suppose, for example, that Jones has killed before, on a number of occasions and in similar circumstances. He isn't such a hardened criminal that he *can't* view his actions from a moral point of view and feel an appropriate revulsion; but the idea that *this person* might be heading straight toward murder, looking neither to the right nor the left, and suddenly, in one simple step, undergo a PAP-relevant shift in attitude, seems like an appeal to *magic*. This is to expect too much of him, and the judgment that he is morally responsible for killing Smith surely does not rest on this expectation being a reasonable one. This attempt to evade the dilemma posed in the first paragraph of this reply must be judged a failure.

Objection 3. If there are no PAP-relevant alternatives once Jones finds himself thinking about killing Smith, there may still be PAP-relevant alternatives available prior to the events described in *Revenge*. In particular, Jones's responsibility for killing Smith might derive from his responsibility for being a certain kind of person (e.g., the kind of person who could decide to kill someone, without even considering doing otherwise, just because he was humiliated by him at a party), which in turn is grounded in earlier character-forming choices which themselves satisfy PAP.²⁶

Reply. This objection fails on two counts. First, it simply challenges the critic of PAP to concoct settings for these earlier character-forming episodes that are themselves structurally similar to *Revenge*, in that a counterfactual intervener is present and the only PAP-relevant alternatives are buffered alternatives. Since the objector's appeal to ancient history is an apparent admission that *Revenge* itself contains no PAP-relevant alternatives accessible to Jones, it's not clear on what grounds the objector might insist that these earlier alternatives could not be rendered similarly inaccessible.

Call this extension of the buffer strategy into Jones's past *Global Revenge*. There is one difficulty with this response that needs to be addressed before moving on to the second reply to the objection.²⁷ For Jones's earlier episodes of moral agency to conform to the pattern of *Revenge*, considering an alternative must always be a necessary condition for Jones to access that alternative, the device must be set to intervene if this necessary condition is satisfied, and the necessary condition must never be satisfied. This means that the Jones of *Global Revenge* apparently *never* considers an alternative to what he actually does, and such a

26. The idea that moral responsibility ultimately rests on earlier character-forming episodes which themselves satisfy PAP is most closely associated with Robert Kane; see especially his *The Significance of Free Will*.

27. I want to thank Gordon Pettit for raising this objection when I presented the first draft of this paper at the University of Notre Dame.

person just seems too *weird* to count with any confidence as a moral agent: always acting on the first thought that pops into one's head is surely a sign of arrested development, neurological disorder, or other defect.²⁸ But it's one thing to recognize a deficiency in this person's agency; it's another thing to say just what is missing.

In *Revenge*, Jones considers, wills, and finally executes the tokening of a specific action-type. Candidates for the latter would include *killing (another human being)*, *killing when one feels like it*, *killing when humiliated*, *killing when humiliated in front of one's friends*, and so on (depending on the specificity of the type); also types in which killing isn't salient, such as *acting on one's immediate impulses without considering the cost to others*. One way to understand the connection between *Revenge* and a putative character-forming episode is to suppose that there is some action-type Θ —one that the objector regards as morally perspicuous—which is tokened on both occasions. It's because of this similarity of type that Jones's responsibility on this earlier occasion can "transfer" to his decision in *Revenge*, despite his inability to access an alternative to this decision. Moreover (the objector might continue), this earlier episode can be genuinely character-forming, entitling it to confer responsibility on later Θ -episodes, only if Jones, on that occasion, considered both Θ and not- Θ , and then decided and acted on his own, without the ministrations of the device. This suggests that the objector's real complaint against *Global Revenge* is something like this:

(G) There is no occasion on which Jones has considered doing Θ and considered not doing Θ and then reached his own decision about what to do, without interference from the device.

This is of course true in *Global Revenge*, since the elimination of PAP-relevant alternatives requires that either Jones's considering Θ or his considering not- Θ serve as a trigger for the device's intervention, and his considering both guarantees its intervention. It isn't possible, then, to respond to (G) by challenging its truth.²⁹ Instead, it will be necessary to argue that Jones might be morally respon-

28. A Thomistic analysis of this strange individual might go like this. He has a *powerful intellect*, inasmuch as he is able to discern the good in any end as soon as he considers it; but he has a *weak will*, inasmuch as his appetite for the good is always moved by the first object proposed to it, before his intellect has a chance to consider alternative ends.

29. It may be possible to challenge (G) by distinguishing two senses in which one might consider a possible action prior to deciding to perform it. One sense is theoretical, which might be rendered this way:

(a) I might do A.

This is a simple recognition that A is among the things that it is possible for me to do. But a second sense is practical:

(b) Would that I might do A?

Here I consider A with an eye toward whether to do it. An affirmative answer to this question might be captured this way:

sible *even if* (G) is true. Here are some examples, of varying value, designed to cast doubt on the idea that (G) undermines Jones's responsibility in *Global Revenge*.³⁰

The Supreme Agent. On some views, at least, God never considers alternatives before acting, though he is the supreme agent.³¹ If this is correct, then agency *per se* does not require a consideration of alternatives at some point in the agent's history, and the burden of proof is on the one who thinks that human agency in particular requires such a consideration. This is all the more true if human agency is most fully agency insofar as it imitates divine agency.³²

Mother Love. A mother who dies while rescuing her newborn child from a burning building is morally praiseworthy for so doing, even (perhaps especially) if she never considers doing otherwise, and the presence of a Frankfurt mechanism, programmed to make her decide to save the child if she considers not doing so, would do nothing to mitigate our praise. If the mother's moral responsibility in this case depends on there being some (earlier) occasion on which she considered doing Θ , considered not doing Θ , and chose between them, what would Θ have to be so that the requirement would be even remotely plausible?

The Rehearsal. When a parent risks his or her life for a child, the parent's action seems *natural* in a way that might not require character-formation. (Perhaps it's acting in the *contrary* manner that would require explanation in terms of earlier character-deformation episodes). But even where character development is relevant and indeed necessary, it isn't always necessary to achieve it through actual episodes of moral choice. Agents sometimes *rehearse* choices for later, in situations in which they are not in fact presented with the choice or even in a position to be presented with it. This happens when agents entertain "what-if" scenarios, reflect on the experiences of role-models who have been in similar situations, read edifying *contes moraux*, listen to sermons on the subject, or take an ethics course.

Tax Evasion. Of the various discrete and detectable events that could serve as triggers for a counterfactual intervener, considering an alternative before undertaking to access it is arguably the one best qualified, in the greatest number and variety of cases, to fill the role of a genuinely necessary condition on doing otherwise. It is for this reason that it makes a good universal tripwire for the alternative-eliminator in *Global Revenge*. But there are also occasions on which it is

(c) Would that I might do A!

If (b) is the necessary condition for (c) that serves as a trigger for the intervener, the agent can consider alternatives in sense (a) without having his freedom overridden. The critic of *Global Revenge*, however, will almost certainly insist that a moral agent must have experience of considering alternate possibilities in sense (b), not just in sense (a).

30. I'm not convinced that employing (G) as an indictment of *Global Revenge* doesn't simply beg the question against PAP's critics; if it does, the four examples that follow may not be necessary.

31. Descartes, for example, maintains in *Meditation IV* that God would have no need to *deliberate*.

32. This paragraph makes no mention of moral responsibility, which would apply to God (if it applies at all) in a very different sense than it applies to human beings. But it's not God's acting without considering the alternative(s) that makes his status with respect to moral responsibility different than that of human beings.

plausible to suppose some further necessary condition for accessing a morally relevant alternative, beyond merely considering the alternative. For example, when a person has some initial inclination to do something, doing otherwise may require not only considering the alternative, but considering some reason for doing otherwise. Or when strongly tempted to do something, resisting the temptation may require not only considering the alternative and having some reason for choosing the alternative, but also attaching sufficient weight to the reason. Derk Pereboom's *Tax Evasion*, in which a man cheats on his taxes and might have refrained from cheating only if a moral reason had occurred to him with sufficient force, is a case of this sort. In such cases, the device may be programmed to intervene when and only when one of these further necessary conditions is satisfied, allowing Jones to consider both Θ and not- Θ without the device automatically interfering in his decision process.³³

So much for the first response to the objection, which takes over the alternative-elimination strategy employed in *Revenge* and endeavors to apply it to earlier episodes that are candidates for character-formation. The second response is that it's not clear that the appeal to earlier alternatives does the job even when such alternatives are available. While character-forming choices undoubtedly play an under-appreciated role in moral responsibility, they explain the wrong thing when it comes to cases like *Revenge* (or, for that matter, *Assassin*). In *Sociopath*, for example, Jones is such that the kind and degree of humiliation he suffered at Smith's hands makes the decision to kill Smith psychologically irresistible. In this case Jones's moral responsibility may piggyback on earlier choices by which he became the kind of person who can't do other than murder Smith. (If there were no earlier choices of the right sort—for example, if Jones's sociopathy were primarily the result of abuse he suffered as a child—this would mitigate Jones's responsibility, as PAP implies.) But *Revenge* is not *Sociopath*; the Jones of *Revenge*, absent the counterfactual intervener, is free to decide or not to decide to kill Smith. Whether Jones is the kind of person who would commit murder when insulted, without giving it a second thought, depends on what he ends up doing, and so long as it is possible for him to do otherwise, it is possible for him to show by his actions that he is *not* that kind of person, despite earlier character-forming choices. *Adding the intervener changes nothing*. Jones now has no morally relevant alternative, but this is owing to the presence of the device, not to the formative power of some earlier choice whose satisfaction of PAP might apply, by extension, to Jones's choice in *Revenge*. The appeal to earlier character-forming choices does not successfully engage the challenge posed by buffer cases.

CONCLUSION

This does not (it goes without saying) dispense with all objections that might be raised against buffer cases; but these seem the most obvious. I conclude that PAP should be rejected. Of course this says nothing about what the right requirements

33. *Living without Free Will*, 18–22.

for moral responsibility might be—even whether those requirements will be compatibilist or incompatibilist.³⁴ What does follow is that alternate possibilities, even if they normally accompany moral responsibility, are no part of what makes an agent morally responsible for what she does.³⁵

34. John Fischer, Mark Ravizza, and Harry Frankfurt himself draw compatibilist morals from the failure of PAP, while Eleonore Stump, Derk Pereboom, Linda Zagzebski, and I reject alternate possibilities while insisting that free and morally responsible agency is incompatible with causal determinism.

35. Versions of this paper were read at the University of Notre Dame, Westmont College, Oxford University, the University of California at Riverside, the University of Minnesota at Morris, and Talbot School of Theology, as well as at the Eastern meeting of the American Philosophical Association. I want to thank the audiences at these venues—especially Gordon Pettit and Cyrille Michon—as well as my commentators (Daniel von Wachter at Oxford, Michael McKenna at the APA).