

ALTERNATIVE POSSIBILITIES AND CAUSAL HISTORIES

Derk Pereboom
University of Vermont

1. Two libertarian attempts to revitalize a principle of alternative possibilities.

Two kinds of objections against Frankfurt-style arguments have recently received significant attention. A first sort, advocated by Robert Kane and systematically formulated by David Widerker, has the following structure. For any Frankfurt-style case, if causal determinism is assumed, the libertarian will not have and cannot be expected to have the intuition that the agent is morally responsible.¹ If, the other hand, libertarian indeterminism is presupposed, an effective Frankfurt-style scenario cannot be devised, for any such case will fall to a dilemma. In Frankfurt-style cases, in the actual situation there is always a prior sign by which the intervener can know that the agent will perform the action she does, and which signals the fact that intervention is not necessary. If in the proposed case the sign causally determines the action, or if it is associated with something that does so, then the intervener's predictive ability can be explained, but then the libertarian would not have the intuition that the agent is morally responsible. If the relationship between the sign and the action is not causally deterministic in such ways, then the libertarian can claim that the agent could have done otherwise despite the occurrence of the prior sign. Either way, some principle of alternative possibilities emerges unscathed.

The second type of objection to Frankfurt-style arguments has been developed by Michael Otsuka, Keith Wyma, and Michael McKenna.² The main idea behind this approach is to propose an alternative-possibilities requirement distinct from the more common sort, one which can withstand arguments of the sort that Frankfurt developed.³ Abstracting away from the details of specific proposals, this strategy claims that for an agent to be morally responsible for an action what is required is that she could have avoided the moral responsibility she actually has for her action, and that, further, no Frankfurt-style case can falsify this requirement. For Frankfurt-style cases all involve some alternative possibility that would serve as the trigger for intervention, and if the interven-

tion had occurred the agent would have avoided the moral responsibility that in the actual situation she has for her action.

I will present a case that defends Frankfurt against the sort of objection that Kane and Widerker raise. This example differs in important respects from others recently devised to answer this objection, and I shall contend that it avoids difficulties that threaten these other cases. Moreover, against the second objection I will argue that any alternative-possibilities condition of the sort that Otsuka, Wyma, and McKenna have proposed that plausibly might explain why an agent is morally responsible falls to this new sort of case as well. (This claim leaves open the prospect, which I acknowledge, that there is an alternative-possibilities condition not relevant to explaining why an agent is morally responsible that is necessary for moral responsibility.)

2. The first objection against Frankfurt-style arguments.

The current discussion of each of the two objections prominently features the notion of a *robust* alternative possibility. I agree with Fischer's claim that for a principle of alternative possibilities to be plausible, the kinds of alternative possibilities that it specifies must be robust, and indeed the notion of robustness plays a prominent part in my own account.⁴ The intuition underlying the alternative-possibilities requirement is that if, for example, an agent is to be blameworthy for an action, it is crucial that she could have done something to avoid being blameworthy—that she could have done something to get herself off the hook. If having an alternative possibility does in fact play a role in explaining an agent's moral responsibility for an action, it would have to be robust in the sense that as a result of securing that alternative possibility, the agent would *thereby* have avoided the responsibility she has for the action she performed—it would be her securing of that alternative possibility *per se* that would explain why the agent would have avoided the responsibility. Following Fischer, here is an example of the non-robust sort of alternative possibility. Imagine a Frankfurt scenario in which Jones very much wants to kill Smith, but it turns out that he will kill Smith only if he in fact blushes beforehand. Then Jones's failure to blush might be the sign that would trigger the intervention of Black, who wants to ensure that Jones kills Smith. Suppose that Jones acts without intervention. Then we might well have the intuition that he is morally responsible, despite the fact that (as it appears) he could not have done otherwise. A libertarian might respond that there was an alternative possibility available to Jones after all: he could have failed to blush. However, failing to blush does not meet our criterion of robustness. For if Jones had failed to blush, he would not *thereby* have avoided responsibility for evading killing Smith—it would not be the failure to blush *per se* that would explain why Jones would not be blameworthy. By typical libertarian intuitions, a robust sort of alternative possibility would involve *willing* to do otherwise than to perform the action the agent in fact wills to perform.⁵

Widerker's version of the first objection has the following structure.⁶ The case at issue is the one we have just encountered, in which Jones wants to kill Smith, but Black is afraid that Jones might become fainthearted, and so he is prepared to intervene if Jones fails to show a sign that he will kill Smith. The sign that he will kill Smith is his blushing at t1. The important features of the scenario are these:

- (1) If Jones is blushing at t1, then, provided no one intervenes, Jones will decide at t2 to kill Smith.
- (2) If Jones is not blushing at t1, then, provided no one intervenes, he will not decide at t2 to kill Smith.
- (3) If Black sees that Jones shows signs that he will not decide at t2 to kill Smith, that is, sees that Jones is *not* blushing at t1, then Black forces Jones to decide at t2 to kill Smith; but if he sees that he is blushing at t1, then he does nothing.

Finally, suppose that Black does not have to show his hand, because

- (4) Jones is blushing at t1, and decides at t2 to kill Smith for reasons of his own.⁷

Although the case is meant to show that Jones is morally responsible despite the fact that he could not have done otherwise, Widerker claims that this conclusion is not forced on the libertarian:

Note that the truth of (1) cannot be grounded in the fact that Jones's blushing at t1 is, in the circumstances, causally sufficient for his decision to kill Smith, or in the fact that it is indicative of a state that is causally sufficient for that decision, since such an assumption would... [not be] accepted by the libertarian. On the other hand, if (1) is not thus grounded, then the following two options are available to the libertarian to resist the contention that Jones's decision to kill Smith is unavoidable. He may either reject (1), claiming that the most that he would be prepared to allow is

- (1a) If Jones is blushing at t1, then Jones will *probably* decide at t2 to kill Smith...

But (1a) is compatible with Jones's having the power to decide not to kill Smith, since there is the possibility of Jones's acting out of character. Or the libertarian may construe (1) as a conditional of freedom in Plantinga's sense...that is, as

- (1b) If Jones is blushing at t1, then Jones will *freely* decide at t2 to kill Smith, [in a sense that allows that the agent could have decided otherwise]⁸

in which case the libertarian may again claim that in the actual situation when Jones is blushing at t1, it is within his power to refrain from deciding to kill Smith at t2.⁹

Widerker's is a very important objection, and it serves as a test for the effectiveness of any Frankfurt-style argument. One point of clarification: If the libertarian that Widerker supposes Frankfurt must convince is simply presupposing a principle of alternative possibilities, then one could not expect that a Frankfurt-style argument would dislodge his view. But Widerker, I think, does not intend that his libertarian simply presuppose this principle, but rather only the claim that moral responsibility is incompatible with an action's having a deterministic causal history. I will proceed with this understanding of Widerker's objection.

3. Problems for recent attempts to answer Widerker.

Several critics have attempted to construct Frankfurt-style arguments that escape this objection. The kinds of cases that have been used in these arguments can be divided into two categories:

- (a) those in which the relationship between the prior sign and the action is causally deterministic, and the indeterminism that makes for the agent's libertarian freedom is present in the causal history of the action before the prior sign

and

- (b) those in which the prior sign is eliminated altogether.

Eleonore Stump and Ishtiyaque Haji have constructed examples in category (a),¹⁰ while David Hunt, and Alfred Mele together with David Robb have devised scenarios in category (b).¹¹

In my view, the cases that have been devised in each of these categories face significant problems. First, (a)-type situations are difficult to construct so that they are effective against Widerker's objection. Stump's and Haji's examples have serious drawbacks. In Stump's case, Grey, the neurosurgeon, wants to ensure that Jones will vote for Reagan. Grey finds that every time Jones decides to vote for Republicans, the decision regularly correlates with *the completion* of a sequence of neural firings in Jones's brain that always includes, near the beginning, the firing of neurons a, b, and c. Jones's deciding to vote for Democratic candidates is correlated with the completion of a neural sequence that always includes, near the beginning, the firing of neurons x, y, and z. Whenever Grey's neuroscope detects the firing of x, y, and z, it disrupts that sequence, with the result that this sequence is not brought to completion. Instead, the device activates a coercive mechanism that makes Jones vote Republican. Crucially, Stump specifies that the firing of x, y, and z does not constitute a decision, and in her view the occurrence of this sequence would not count as a robust alternative possibility. If, on the other hand, the neuroscope detects the firing of a, b, and c, it allows the sequence to proceed to completion and the

decision to vote Republican to occur.¹² Stump specifies that the decision is indeed a causal outcome of the neural sequence.¹³ What makes the agent libertarian is that the neural sequence is not the outcome of a causal chain that originates in a cause outside him. Rather, it is the outcome of a causal chain that originates, at least to a significant extent, *in an act of the agent which is not the outcome of a causal chain that originates in a cause outside the agent*. Here Stump suggests the Aquinas-inspired view that the neural sequence is the outcome of a causal chain that originates in the agent's intellect and will.¹⁴

But as Stewart Goetz points out, to assess this case, one needs to know more about the psychological features of the act performed by the agent to cause the neural process. If this act is causally determined, then Stump's agent would appear not to be free in the libertarian sense. But if it is not causally determined, then he might well have robust alternative possibilities for action. If it is the act of intending to make a decision, for example, and if the indeterminism of that act allows for the agent also to have avoided intending to make the decision, then the case might well include a robust alternative possibility after all.¹⁵ Note that in Stump's setup, the agent's performance of that act—which constitutes the agent's crucial libertarian causal role—precedes the possible intervention.¹⁶

More generally, the challenge for Stump is to characterize the agent's causal role so that (i) her action is not causally determined (by factors beyond her control) and (ii) her action does not involve robust alternative possibilities. A case of the sort that Stump devises is subject to the following dilemma: If the indeterminism (whether or not it is a characteristic of the sort of agent's act she has in mind) that occurs prior to the neural sequence is significant enough to make the action a libertarian freely-willed action, then it has not been ruled out that the indeterministic juncture features a robust alternative possibility. If Stump were to reject the claim that there is a robust alternative possibility at this point, then it would remain open to a libertarian (like Widerker) to deny that the agent has genuine libertarian free will. Perhaps it is possible to embellish Stump's example to answer this objection. But to me it is not clear that there could be a plausible Frankfurt-style case in which the action is not causally determined by factors beyond the agent's control (in a way that would satisfy the libertarian) and she lacks robust alternative possibilities if the intervention would occur after the crucial indeterministic juncture.

In Haji's example, the sort of libertarian agency attributed to the agent consists in its being undetermined which of various considerations will enter the mind of the agent in deliberation. So, at the outset of Jones's deliberation it is causally open whether he will kill Smith, because it is causally undetermined whether various considerations will enter his mind at the onset of his musings. The infallible predictor, Black,

intervenes if and only if he believes that Jones will not make the decisive best judgment that favors the decision to kill Smith which he, Black, wants Jones to

make. Specifically, should Jones make the judgment that he ought not to kill Smith, *then* (and only then) will Black intervene and cause Jones to alter the judgment.¹⁷

There is no need for Black to intervene “as Jones decides appropriately on his own,” and one’s intuition is that Jones could be morally responsible for his decision. However, one problem for Haji’s case is that it was open to Jones to have made the decisive best judgment that he ought not to kill Smith, and this alternative possibility seems robust. For it appears plausible that if Jones had made the decisive best judgment that he ought not to kill Smith he would thereby have avoided the responsibility he has for the action he actually performed—it would be his securing of this alternative possibility *per se* that would explain why he would have avoided this responsibility. Another difficulty for Haji’s example is that many libertarians would not allow that the sort of indeterminacy he specifies could be significant for moral responsibility because it fails to provide the agent with enhanced control. A case in which the relevant considerations indeterministically enter the mind of the agent, whereupon his judgment and decision are determined, would seem to exhibit no more control by him than a situation in which such considerations deterministically enter the mind of the agent, whereupon the agent’s judgment and decision are causally determined.¹⁸

4. Frankfurt-style scenarios without prior signs.

Cases in category (b) exemplify a different of strategy for opposing alternative-possibility conditions. In these cases there are no prior signs to guide intervention, not even non-robust flickers of freedom. One ingenious scenario in this category is presented by Mele and Robb.¹⁹ The example features Bob, who inhabits a world in which determinism is false:

At t1, Black initiates a certain deterministic process P in Bob’s brain with the intention of thereby causing Bob to decide at t2 (an hour later, say) to steal Ann’s car. The process, which is screened off from Bob’s consciousness, will deterministically culminate in Bob’s deciding at t2 to steal Ann’s car unless he decides on his own to steal it or is incapable at t2 of making a decision (because, e.g., he is dead at t2)... . The process is in no way sensitive to any “sign” of what Bob will decide. As it happens, at t2 Bob decides on his own to steal the car, on the basis of his own indeterministic deliberation about whether to steal it, and his decision has no deterministic cause. But if he had not just then decided on his own to steal it, P would have deterministically issued, at t2, in his deciding to steal it. Rest assured that P in no way influences the indeterministic decision-making process that actually issues in Bob’s decision.

Mele and Robb claim that Bob is plausibly morally responsible for his decision. I think that their argument may in fact be successful, but that their development of the case raises one problem that could undermine it. Mele and Robb

discuss several potential problems for their scenario, one of which is whether we can make sense of what would happen at t_2 if P and Bob's indeterministic deliberative process were to diverge at t_2 . Here is how they handle the difficulty:

The issue may be pictured, fancifully, as follows.²⁰ Two different "decision nodes" in Bob's brain are directly relevant. The "lighting up" of node N1 represents his deciding to steal the car, and the "lighting up" of node N2 represents his deciding *not* to steal the car. Under normal circumstances and in the absence of preemption, a process's "hitting" a decision node in Bob "lights up" that node. If it were to be the case both that P hits N1 at t_2 and that x does not hit N1 at t_2 , then P would light up N1. If both processes were to hit N1 at t_2 , Bob's indeterministic deliberative process, x, would light up N1 and P would not. The present question is this. What would happen if, at t_2 , P were to hit N1 and x were to hit N2? That is, what would happen if the two processes were to "diverge" in this way? And *why*?

We extend Bob's story as follows. Although if both processes were to hit N1 at t_2 , Bob's indeterministic deliberative process, x, would preempt P and light up N1, it is also the case that if, at t_2 , P were to hit N1 and x were to hit N2, P would prevail. In the latter case, P would light up N1 and the indeterministic process would not light up N2. Of course, readers would like a story about why it is that although x would preempt P in the former situation, P would prevail over x in the latter. Here is one story. By t_2 , P has "neutralized" N2 (but without affecting what goes on in x). That is why, if x were to hit N2 at t_2 , N2 would not light up.²¹ More fully, by t_2 P has neutralized all of the nodes in Bob for decisions that are contrary to a decision at t_2 to steal Ann's car (e.g., a decision at t_2 not to steal anyone's car and a decision at t_2 never to steal anything).²² In convenient shorthand, by t_2 P has neutralized N2 and all its "cognate decision nodes." Bear in mind that all we need is a conceptually possible scenario, and this certainly looks like one.²³

The aspect of this story that might raise the libertarian's eyebrows is P's neutralization of N2 and all its cognate decision nodes. For he might be tempted to claim that P's neutralizing procedure is equivalent to P's causal determination of Bob's decision to steal the car. On the other hand, Mele and Robb do specify that P's neutralizing activity does not affect what goes on in Bob's indeterministic decision making process, and so it would seem that P would not causally determine the decision. How can we shed light on this difficulty?

Let us examine an approach in category (b) that more vigorously exploits the neutralization idea. A strategy of this type has become known as "blockage," and has been developed by David Hunt.²⁴ Here is a way of presenting this sort of approach that I think is especially powerful. Consider two situations.

Situation A: Ms. Scarlet deliberately chooses to kill Colonel Mustard at t_1 , and there are no factors beyond her control that deterministically produce her choice. When she chooses to kill the Colonel, she could have chosen not to kill him. There are no causal factors that would prevent her from not making the choice to kill Colonel Mustard.

In these circumstances, Ms. Scarlet could be morally responsible for her choice. But then, against an alternative-possibilities principle one might employ a counterfactual version of this situation:

Situation B: Ms. Scarlet's choice to kill Colonel Mustard has precisely the same actual causal history as in A. But before she even started to think about killing Colonel Mustard, a neurophysiologist had blocked all the neural pathways not used in Situation A, so that no neural pathway other than the one employed in that situation could be used. Let us suppose that it is causally determined that she remain a living agent, and if she remains a living agent, some neural pathway has to be used. Thus every alternative for Ms. Scarlet is blocked except the one that realizes her choice to kill the Colonel. But the blockage does not affect the actual causal history of Ms. Scarlet's choice, because the blocked pathways would have remained dormant.

One might, at least initially, have the intuition that Ms. Scarlet could be morally responsible for her choice in B as well. Yet for an incompatibilist this intuition might well be undermined upon more careful reflection on whether in B Ms. Scarlet retains libertarian freedom. One important question about such blockage cases is one Fischer asks: *Could* neural events bump up against, so to speak, the blockage?²⁵ If so, there still may be alternative possibilities for the agent. But if not, why aren't the neural events causally determined partly by virtue of the blockage? Kane's response to blockage cases reflects this worry:

In [a case in which every other alternative is blocked except the agent's making A at t], of course, there *are* no alternative possibilities left to the agent; every one is blocked except the agent's choosing A at t. But now we seem to have determinism pure and simple. By implanting the mechanism in this fashion, a controller would have predetermined exactly what the agent would do (and when); and, as a consequence, the controller, not the agent, would be ultimately responsible for the outcome. Blockage by a controller that rules out all relevant alternative possibilities is simply predestination; and on my view at least, predestination runs afoul of ultimate responsibility.²⁶

In response, one might claim that in the standard Frankfurt-style cases the relevant action is inevitable, but the intuition that the agent is morally responsible for it depends on the fact that it does not have an actual causal history by means of which it is made inevitable. What makes the action inevitable is rather some fact about the situation that is not a feature of its actual causal history, and hence, the action's being inevitable need not make it the case that it is causally determined. But how is the blockage case different from the standard Frankfurt-style cases? After all, the blockage does not seem to affect the actual causal history of the action.

Nevertheless, perhaps Kane's response can be defended. Two-situation cases of the above sort might be misleading just because it is natural to assume that the actual causal history of an event is essentially the same in each, given that the only difference between them is a restriction that would seem to have no actual effect on the event. But consider a simple two-situation case modelled on a reflection of Hunt's.²⁷ Imagine a universe correctly described by Epicurean physics: what exists at the most fundamental level is atoms and the frictionless void, and there is a determinate downwards direction in which all atoms naturally fall—except if they undergo uncaused swerves.

Situation C: A spherical atom is falling downward through space, with a certain velocity and acceleration. Its actual causal history is indeterministic because at any time the atom can be subject to an uncaused swerve. Suppose that the atom can swerve in any direction other than upwards. In actual fact, from t_1 to t_2 it does not swerve.

A counterfactual situation diverges from C only by virtue of a device that eliminates alternative possibilities and all differences thereby entailed:

Situation D: The case is the identical to C, except that the atom is falling downward through a straight and vertically oriented tube whose interior surface is made of frictionless material, and whose interior is precisely wide enough to accommodate the atom. The atom would not have swerved during this time interval, and the trajectory, velocity, and acceleration of the atom from t_1 to t_2 are precisely what they are in C.

One might initially have the intuition that the causal history of the atom from t_1 to t_2 in these two situations is in essence the same. However, this intuition could be challenged by the fact that the restrictions present in D but not in C may change this causal history from one that is essentially indeterministic to one that is essentially deterministic. For since the tube prevents any alternative motion, it would seem that it precludes any indeterminism in the atom's causal history from t_1 to t_2 . And if the tube precludes indeterminism in this causal history, it would appear to make the causal history deterministic. Whether this line of argument is plausible is difficult to ascertain, but neither is it obviously implausible.

This problem could make it hard to assess moral responsibility in blockage cases. 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.

5. A new Frankfurt-style scenario.

Here is my case:

Tax evasion, Part 1: Joe is considering whether to claim a tax deduction for the substantial local registration fee that he paid when he bought a house. He knows that claiming the deduction is illegal, that he probably won't be caught, and that if he is, he can convincingly plead ignorance. Suppose he has a very powerful but not always overriding desire to advance his self-interest no matter what the cost to others, and no matter whether advancing his self-interest involves illegal activity. Furthermore, he is a libertarian free agent. But his psychology is such that the only way that in this situation he could fail to choose to evade taxes is for moral reasons. His psychology is not, for example, such that he could fail to choose to evade taxes for no reason or simply on a whim. In fact, it is causally necessary for his failing to choose to evade taxes in this situation that a moral reason occur to him with a certain force. A moral reason can occur to him with that force either involuntarily or as a result of his voluntary activity (e.g. by his willing to consider it, or by his seeking out a vivid presentation of such a reason). But a moral reason occurring to him with such force is not causally sufficient for his failing to choose to evade taxes. If a moral reason were to occur to him with that force, Joe could, with his libertarian free will, either choose to act on it or act against it (without the intervener's device in place). But to ensure that he decide to evade taxes, a neuroscientist now implants a device which, were it to sense a moral reason occurring with the specified force, would electronically stimulate his brain so that he would decide to evade taxes. In actual fact, no moral reason occurs to him with such force, and he chooses to evade taxes while the device remains idle.

In this situation, Joe could be morally responsible for deciding to evade taxes despite the fact that he could not have chosen otherwise. The prior sign does not causally determine his decision. There is indeed an alternative possibility, that a moral reason occur to him with a certain force. But such a possibility is insufficiently robust to ground his moral responsibility for tax evasion. For, again, the deeper intuition underlying alternative-possibilities requirement is that if, for example, an agent is to be blameworthy for an action, it is crucial that

she could have done something to avoid this blameworthiness. If alternative possibilities were to play a role in explaining an agent's moral responsibility for an action (in a way independent of an intuition about its actual causal history), it would be because as a result of securing an alternative possibility instead, he would thereby have avoided the responsibility he has for the action he performed. However, if Joe had made a reason for an alternative action occur to him with a certain force, he would not *thereby* have avoided responsibility for evading taxes. For his making the reason for an alternative action occur to him is compatible with his never deciding to perform the alternative action, or even ever being inclined to perform that action, and choosing to evade taxes instead.

This example fits neither description (a) or (b). Rather, it is a case that has the following features:

- (i) The agent clearly has free will according to most libertarian views.
- (ii) What would trigger the intervention is a "flicker" that is insufficiently robust to explain the agent's moral responsibility for the decision in question.
- (iii) It does not ground the truth of the analogue of Widerker's

(1) If Jones is blushing at t1, then, provided no one intervenes, Jones will decide at t2 to kill Smith,

which is

(1') If a moral reason does not occur to Joe with a certain force, then, provided no one intervenes, Joe will decide to evade taxes,

in causal determinism, while at the same time not endorsing the analogues of Widerker's

(1a) If Jones is blushing at t1, then Jones will [only] *probably* decide at t2 to kill Smith,

and

(1b) If Jones is blushing at t1, then Jones will *freely* decide at t2 to kill Smith, [in a sense that allows that the agent could have decided otherwise],

which are

(1a') If a moral reason does not occur to Joe with a certain force, then he will [only] probably decide to evade taxes,

and

(1b') If a moral reason does not occur to Joe with a certain force, then he will freely decide to evade taxes (in a sense that allows that the agent could have decided otherwise).

The absence of what would trigger the intervention at some particular time (the role of this absence is played by Jones's blushing at t_1 in Widerker's case, and by the non-occurrence of a moral reason with the requisite force at some particular time in Tax Evasion), or a state indicated by this absence, will not together with all the other actual facts about the situation, causally determine the decision. Joe's decision is not causally determined by the non-occurrence at any particular time of a moral reason with sufficient force. For at any point in the causal history of the action prior to the choice a moral reason could have occurred to him with sufficient force, even as a result of his own (undetermined) voluntary activity. But (*contra* 1a) the decision *will* occur (and not just probably occur) in the absence of what would trigger the intervention, even though it is not causally determined, because what would trigger the intervention is causally necessary (but not causally sufficient—thus not causally determining) for the decision's not occurring. Hence, (*contra* 1b) there is a libertarian sense in which the agent can *freely* decide to perform the action, but *without* its being the case that she could have decided otherwise.

Seeing how this example responds to Kane's version of the objection highlights the value of having the cue for intervention be causally necessary but not sufficient for the action, while ensuring that up to the time of the decision itself, the agent is not causally determined to make it. Kane argues, first of all, that supposing the case is to convince the libertarian, then if the agent in the example decides on his own, then this decision must be causally undetermined. Now if the intervention does occur, then the agent is not morally responsible. But if the neuroscientist "does not intervene to predetermine the outcome and the indeterminacy remains in place until the choice is made—so that the outcome is [a "self-forming willing"]—then the agent...is ultimately responsible for it. However, then it is also the case that the agent *could have done otherwise*."²⁸ However, let the cue for intervention be a causally necessary condition of the alternative, such as, in our example, the occurrence to the agent of a moral reason with a certain force. Then if the neuroscientist does not intervene, even though the indeterminacy remains in place until the choice is made, *it is not the case*, contrary to Kane's supposition, that the agent could have decided or could have done otherwise. For in order to decide otherwise, a moral reason would have had to occur with the requisite force, and then the device would have been activated.

One might reply that in order for Joe to be responsible for his action, his moral psychology must have been set up by crucial choices of his for which there were robust alternative possibilities.²⁹ But to see that this sort of answer is mistaken, consider:

Tax Evasion, Part 2: Joe was raised in a context in which people are typically self-interested in the sort of way he is now. His parents, for example, had this sort of psychological profile. But he was also raised to reflect critically on his values as soon as he was able. Like most of us, he initially

accepted his family's values, and he held them very strongly. Joe then learned about competing positions, but upon serious reflection, he rejected them. Suppose that for him to abandon his initial moral view it was causally necessary that a reason for accepting a competing position occur to him with a certain force, and this could occur either involuntarily or as a result of his voluntary activity. Were such a reason to occur to him with that force, he could still choose to retain or reject his values (in the absence of the neuroscientist's device). But the neuroscientist, knowing all of this about his psychology, sets up his device before Joe begins critical reflection on his moral views, so that if a reason to accept a competing view were to occur to him with the specified force, it would electronically stimulate his brain to retain his initial moral position. But in actual fact, the device remains inert, for although he considers reasons to accept competing views, these reasons never occur to him with sufficient force to trigger the device.

Thus, although Joe's moral psychology was not set up by crucial choices of his for which there were robust alternative possibilities, he could still be morally responsible for evading taxes.

Consequently, this type of objection can be answered. Even presupposing libertarianism, we have not yet encountered a principle of alternative possibilities that plausibly has a significant role in grounding moral responsibility.

6. Answering the second objection.

As we have seen, the second type of attack on Frankfurt-style strategies has been advanced by Michael Otsuka, Keith Wyma, Michael McKenna.³⁰ Otsuka claims that a necessary condition on blameworthiness for an act of a given type is that the agent, as a result of his voluntary endeavor, could have instead behaved in a manner for which he would have been entirely blameless.³¹ Wyma, in a similar vein, claims that a person is morally responsible for something she has done, A, only if she has failed to do something she could have done, B, such that doing B would have rendered her not morally responsible for A. McKenna argues that in all of the successful Frankfurt-style cases, the agent has the power either to be the author of his action or not, and that it is precisely this sort of alternative possibility that is significant for moral responsibility.

These philosophers assume that their conditions could not be falsified by Frankfurt cases because the successful versions involve the possibility that the agent could have voluntarily done, or been the author of, something that would have triggered the intervention, whereupon she would not have been morally responsible for the act in question. Fischer argues that these conditions could be undermined by cases in which intervention would be triggered by an involuntary flicker, in which the agent does not have a voluntary alternative possibility.³² However, in setting up an involuntary flicker case, one must be careful

to avoid the problem that Kane and Widerker raise by ensuring that Joe remains a libertarian free agent. We might try to do so by changing Joe's psychology in Part 1 so that in his situation a vivid presentation by an external source would now be required for a moral reason to occur to him with the requisite force. Given the strength of his self-interest and the level of his commitment to morality, since the benefit to himself at stake is significant and the damage to others that would result is not especially great, he could not make a moral reason occur to him with the right force voluntarily, and he could not voluntarily seek the sort of external presentation of moral reasons that would make them occur to him with this force. The relevant facts about the history of his psychology are given by Tax Evasion, Part 2, with the addition that he could not make a moral reason occur to him with sufficient force because of the strength of his self-interest and the weakness of his moral commitment, and the extent to which these features of his psychology have become ingrained. This specification is psychologically plausible—there is much that typical agents could not bring themselves to do because of their commitments. But again (without the intervener's device in place), if a moral reason were to occur to him with the specified force, Joe could, with the power of his libertarian free will, either decide to act on it or act against it. The neuroscientist's device is set up so that it senses when a moral reason occurs to him with such force, and he decides without the device intervening. If Joe could be morally responsible for deciding to evade taxes in this case, then it would directly undermine Otsuka's condition, since Joe could not, *as a result of his voluntary endeavor*, have behaved in such a manner for which he would not have been entirely blameless. It would show McKenna's to be mistaken, for Joe lacks the power *not to be the author of* the tax evasion, and similarly for Wyma's, since it is not the case that he failed to do something he *could have done*, such that doing it would have rendered him not morally responsible for evading taxes.

But the advocate of an alternative-possibilities requirement might argue that if Joe were morally responsible despite his not meeting these conditions it would be because his not doing so is explained by his moral psychology, which, in turn, results from certain crucial choices that he made that do fulfil conditions of the general sort we are now examining. Joe's inability to make a moral reason occur to him with sufficient force is explained by the fact that he chose to retain the self-interested moral conception that he was raised to hold, and although he could not have chosen otherwise, he failed to do something he could have done (i.e., have a reason for accepting a competing moral position occur to him with a certain force), such that doing so would have rendered him not morally responsible for making this choice (because then the device would have been activated). This claim suggests the following alternative possibility principle:

- (Z) An agent is morally responsible for something she has done, A, only if she has failed to do something she could have done, B, such that

doing B would have rendered her not morally responsible for A, or if she could not have done something that would have rendered her not morally responsible for A, this fact is explained by choices this agent made, $C_{1\dots n}$, which are such that at the time she made them she could have done something, $D_{1\dots n}$, such that doing $D_{1\dots n}$ would have rendered her not morally responsible for $C_{1\dots n}$.

It is not obvious to me that this principle has a counterexample, and that therefore (Z) could be an alternative-possibilities condition that is necessary for moral responsibility. We might revise Part 2 of Tax Evasion so that for Joe to have abandoned his initial moral view it was causally necessary that a reason for accepting a competing position occur to him with sufficient force, but that such a reason could occur to him only involuntarily, and never as a result of his voluntary activity. However, then we would want to know why his psychology has this feature, and the sort of explanation that suggests itself is that he has been so thoroughly indoctrinated by his upbringing that his ability to evaluate his moral view rationally has been impaired by factors beyond his control. As a result, there is a strong pull to the claim that Joe is not morally responsible in this situation. There may be another way to construct a counterexample to this principle, but it is not obvious to me how this might be done.

I favor a different kind of objection to the sort of condition that McKenna, Wyma, and Otsuka advocate. First of all, it may be that a condition is necessary for some phenomenon A but sometimes holds by virtue of features that are irrelevant to explaining the nature of A, that do not illuminate the nature of A. I believe that even if it turns out that conditions of the sort at issue are necessary for moral responsibility, they can hold by virtue of features of a situation that are explanatorily irrelevant to what would make an agent morally responsible, and that as a result the condition at issue fails to illuminate the nature of the phenomenon.

Consider Wyma's view, according to which an agent is morally responsible for something she has done, A, only if she has failed to do something she could have done, B, such that doing B would have rendered her morally non-responsible for A. Suppose that Joe could have voluntarily taken a sip from his coffee cup prior to his deciding to take the illegal deduction, not understanding that this action would preclude his evading taxes, because unbeknownst to him, taking the sip would have triggered a bomb that would have killed him. In this situation, he could have behaved voluntarily in such a manner that would have precluded the action for which he was in fact blameworthy, as a result of which he would have been morally non-responsible for this action. But whether he could have voluntarily taken the sip from the coffee cup, not understanding that it would render him blameless in this way, is intuitively irrelevant to whether he is morally responsible for tax evasion. We might say that despite the fact that Joe could have voluntarily taken a sip from his coffee cup, and doing so would have rendered him morally non-responsible for evading taxes, this alter-

native possibility is nevertheless insufficiently robust to have an important role in grounding moral responsibility. Because this sort of alternative possibility would render Wyma's proviso satisfied, conditions of this sort, even if they are necessary for moral responsibility, would appear not to have a significant role in explaining its nature. (In fairness to Otsuka, the problem fails to undermine the sort of position he actually holds, since he resists the claim that an alternative possibilities condition would explain why an agent is morally responsible for an action.)

A condition more plausibly relevant to explaining moral responsibility for an action A is that the agent could have willed otherwise in the following more robust sense: she could have willed to refrain from doing A, or she could have willed to perform an action that she understands would preclude her doing A (or that she understands would preclude her performing an action of A's general sort). However, Tax Evasion does have the right characteristics to undermine this condition, for it shows that Joe could be morally responsible despite the fact that this condition is not satisfied. True, Joe could have willed that a moral reason not to evade taxes occur to him with a particular force. But if Joe were voluntarily to make such a moral reason occur to him with this force, the intervention would take place. He could not then have willed to refrain from evading taxes, or willed to perform an action that he understood would preclude his evading taxes. Nevertheless, Joe is morally responsible for his action. Consequently, if Wyma's condition is refined in this way, so as to eliminate the factors that make it hold for a situation due to features irrelevant to explaining moral responsibility, then it can be shown false. The same type of point can be made about Otsuka's and McKenna's provisos. Conditions of this sort seem to purchase technical indefeasibility at the cost of explanatory relevance to moral responsibility.

7. Causal history incompatibilism.

For the record, I do not intend my arguments to support a compatibilist position. Rather, I endorse incompatibilism, but I believe that the most plausible and fundamentally explanatory incompatibilist principles concern the causal history of an action, and not alternative possibilities. For a powerful and common incompatibilist intuition is that if all of our behavior was "in the cards" before we were born, in the sense that things happened before we came to exist that, by way of a deterministic causal process, inevitably result in our behavior, then we cannot legitimately be blamed for our wrongdoing. By this intuition, if causal factors existed before a criminal was born that, by way of a deterministic process, inevitably issue in his act of murder, then he cannot legitimately be blamed for his action. If all of our actions had this type of causal history, then it would seem that we lack the kind of control over our actions that moral responsibility requires. This intuition does not involve a claim about alternative possibilities. Accordingly, I favor *causal history incompatibilism* over *leeway incompatibilism*. However, I will not develop my positive proposal here.³³

One might place incompatibilists on a continuum ranging from those who hold that moral responsibility requires that an agent could have done or chosen otherwise, to the most radical causal history incompatibilists, who maintain that even in a blockage case an agent can be morally responsible so long as the actual causal history of the action has the right features. Advocates of an alternative-possibilities requirement would tend to advocate an indeterminist condition on the actual causal history of an action only because it would make the existence of alternative possibilities a prerequisite for moral responsibility, for it is a condition about alternative possibilities that is most significant for explaining why agents would be morally responsible. My own incompatibilist view is that moral responsibility requires that the action's actual causal history have certain indeterministic features, but that it might well be that alternative possibilities—not necessarily of the robust sort—are entailed by the actual causal history having these features. Nevertheless, the aspect of the action that has the important role in explaining why agents would be morally responsible is the nature of the actual causal history, and not the alternative possibilities.³⁴

Notes

1. Robert Kane, *Free Will and Values*, (Albany: SUNY Press, 1985), p. 51 n. 25, and *The Significance of Free Will* (New York: Oxford University Press, 1996), pp. 142–4, 191–2; David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104 (1995), pp. 247–61; see also Keith D. Wyma, "Moral Responsibility and Leeway for Action," *American Philosophical Quarterly* 34 (1997), pp. 57–70. Fischer provides a clear and helpful account of these views in "Recent Work on Moral Responsibility," forthcoming in *Ethics*.
2. Michael Otsuka, "Incompatibilism and the Avoidability of Blame," *Ethics* 108 (1998), pp. 685–701; Keith D. Wyma, "Moral Responsibility and Leeway for Action;" Michael McKenna, "Alternative Possibilities and the Failure of the Counterexample Strategy," *Journal of Social Philosophy* 28 (1997), pp. 71–85. Fischer provides a valuable account of these positions in "Recent Work on Moral Responsibility."
3. Examples of more common principles of alternative possibilities are:

An action is free in the sense required for moral responsibility only if the agent could have done otherwise than she actually did.

and

An action is free in the sense required for moral responsibility only if the agent could have chosen otherwise than she actually did.

4. Fischer, *The Metaphysics of Free Will*, (Oxford: Blackwell, 1994), pp. 131–159.
5. See also Mele's characterization of robustness, which I endorse, in "Soft Libertarianism and Frankfurt-Style Scenarios," *Philosophical Topics* 24 (1996), pp. 123–141, at pp. 126–7.
6. cf. Ishtiyaque Haji, *Moral Appraisability* (New York: Oxford University Press, 1998), pp. 34–5.

7. Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," pp. 249–50.
8. The bracketed phrase does not occur in Widerker's text, but it clearly expresses his meaning.
9. Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," p. 250.
10. Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," in *Faith, Freedom, and Rationality*, Jeff Jordan and Daniel Howard Snyder, ed. (Lanham, MD, Rowman and Littlefield, 1996), pp. 73–88;" Haji, *Moral Appraisability*, p. 36.
11. Alfred Mele and David Robb, "Rescuing Frankfurt-Style Cases," *The Philosophical Review* 107 (January 1998), pp. 97–112; David Hunt, "Moral Responsibility and Unavoidable Action," forthcoming in *Philosophical Studies*, cited by Mele and Robb and by Fischer in "Recent Work on Moral Responsibility".
12. Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," pp. 77–8.
13. Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," pp. 79.
14. Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," pp. 80–5.
15. Stewart Goetz, "Stumping for Widerker," *Faith and Philosophy* 16 (1999), pp. 83–9. In this article he develops this and other criticisms in further detail.
16. Stump replies to Goetz's objection in "Dust, Determinism and Frankfurt: A Reply to Goetz," in *Faith and Philosophy* 16 (1999), pp. 413–22, but in my view, she does not lay to rest the worry I just described.
17. Haji, *Moral Appraisability*, p. 36. Haji attributes the inspiration for this sort of libertarianism to Mele, in *Autonomous Agents*, (New York: Oxford University Press, 1995), p. 216. This kind of libertarianism is also suggested by Daniel Dennett, in "Giving Libertarians What They Say They Want," in his *Brainstorms* (Montgomery, VT: Bradford Books, 1978), at p. 295.
18. See, for example, Randolph Clarke, "Agent Causation and Event Causation in the Production of Free Action," *Free Will*, Derk Pereboom, ed. (Indianapolis: Hackett Publishing Company, 1997), p. 286, (a longer version of this article appears in *Philosophical Topics* 24, 1996, pp. 19–48).
19. Mele and Robb, "Rescuing Frankfurt-Style Cases."
20. (Mele and Robb's note.) The picture obviously is neuro-fictional, but it is useful nonetheless.
21. (Mele and Robb's note.) What would happen if Bob's indeterministic deliberative process were to hit N2 at some time t_n prior to t_2 ? In one version of the story, N2 would light up at t_n —Bob would decide at t_n not to steal the car—but then at t_2 , when P hits N1, Bob would change his mind and decide to steal it. In another version—the one we prefer, owing to its relative simplicity—P neutralizes N2 as soon as Black initiates P.
22. (Mele and Robb's note.) David Hunt independently makes a similar suggestion in a forthcoming article (Hunt, ["Moral Responsibility and Unavoidable Action"]).
23. Mele and Robb, "Rescuing Frankfurt-Style Cases," pp. 104–5.
24. Fischer, contribution to a symposium on *The Significance of Free Will*, in *Philosophy and Phenomenological Research*, forthcoming, 1999.
25. Fischer, "Recent Work on Moral Responsibility."
26. Kane, in his reply to Fischer in a symposium on *The Significance of Free Will*, forthcoming in *Philosophy and Phenomenological Research*.
27. From Hunt's personal correspondence with Fischer, cited in Fischer's "Recent Work on Moral Responsibility."

28. Kane, *The Significance of Free Will*, p. 142.
29. Kane suggests a reply of this sort to a similar case, in correspondence.
30. Michael Otsuka, "Incompatibilism and the Avoidability of Blame;" Keith D. Wyma, "Moral Responsibility and Leeway for Action;" Michael McKenna, "Alternative Possibilities and the Failure of the Counterexample Strategy," *Journal of Social Philosophy*, pp. 71–85. Fischer provides a helpful account of these positions in "Recent Work on Moral Responsibility."
31. By way of clarification, Otsuka states:

When I say that one could instead have behaved in a manner for which one would have been entirely blameless, I mean that it was within one's voluntary control whether or not one ended up behaving that way. But I need not claim that the behavior itself must have been voluntary ("Incompatibilism and the Avoidability of Blame," pp. 688–9)

32. Fischer, "Recent Work on Moral Responsibility."
33. I argue for this view in "Determinism *Al Dente*," *Noûs* 29, March 1995, pp. 21–45. For similar positions see Robert Heinaman, "Incompatibilism without the Principle of Alternative Possibilities," *Australasian Journal of Philosophy* 64 (1986), pp. 266–76, and Eleonore Stump, "Libertarian Freedom and the Principle of Alternative Possibilities."
34. I wish to thank David Christensen, John Fischer, Hilary Kornblith, Alfred Mele, and Linda Zagzebski for helpful comments and conversations.