
CAUSATION AND FREEDOM

Author(s): CAROLINA SARTORIO

Source: *The Journal of Philosophy*, NOVEMBER 2012, Vol. 109, No. 11 (NOVEMBER 2012), pp. 629-651

Published by: Journal of Philosophy, Inc.

Stable URL: <https://www.jstor.org/stable/43820735>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*

JSTOR

THE JOURNAL OF PHILOSOPHY

VOLUME CIX, NO. 11, NOVEMBER 2012

CAUSATION AND FREEDOM*

The concept of causation usually plays an important role in the formulation of the problem of freedom and determinism. Despite this fact, and aside from the debate over whether the mysterious relation of “agent-causation” is possible, the literature on that problem very rarely engages in metaphysical debates about the nature of causation or attempts to draw on the metaphysics of causation *per se*. For example, philosophers do not tend to think that the question of whether causation (ordinary, “event-causation”; henceforth, just *causation*) is reducible to counterfactual dependence, or the instantiation of some regularity, or a primitive relation not reducible to other more basic metaphysical concepts, has any bearing on the question of whether freedom is compatible with determinism.¹ Similarly, philosophers do not tend to think that more specific issues in the metaphysics of causation such as the question about the nature of the causal relata, or whether absence causation is possible, or whether causation is transitive, are relevant to this debate. Here I will argue that some debates in the metaphysics of causation actually have a significant bearing on that debate. My main focus will be a popular view in that debate: the *alternative-possibilities* view of freedom and responsibility. I will argue that the metaphysics of

* Thanks to Michael Bergmann, Randolph Clarke, Juan Comesaña, Terry Horgan, Uriah Kriegel, David Owen, and audiences at Arizona State University, Ohio State University, Purdue University, and the University of Arizona. I am especially grateful to Michael McKenna, who provided invaluable comments on several drafts.

¹ In their “Humean Compatibilism” (*Mind*, cxi, 442 (April 2002): 201–23), Helen Beebe and Alfred Mele argue that a certain reductive view about the *laws of nature* (Humeanism) can be used to disarm an important argument for incompatibilism about determinism and freedom. But they do not comment on the relevance of the metaphysics of causation.

0022-362X/12/0911/629–51

© 2012 The Journal of Philosophy, Inc.

629

causation plays an important role in the formulation of the best version of this view.²

I

According to the alternative-possibilities view of freedom and responsibility, the kind of freedom necessary for responsibility (and a kind of freedom that according to some philosophers is incompatible with the truth of determinism) requires acting from a number of “forking paths.” An agent is free in taking a certain path only if he had the ability to take some alternative path. Now, as it stands, this is just a colorful metaphor. How exactly should we understand the forking-paths metaphor? What, in particular, should we take the forking paths to be?

At least in the case of an agent’s responsibility for his own acts, the metaphor has by now acquired something close to a standard interpretation: it is typically cashed out in terms of the famous “Principle of Alternative Possibilities,” which states that a person is morally responsible for having performed a certain act only if he could have done otherwise.³ Now, as it has been pointed out, there are different things that “could have done otherwise” could mean. It could mean, for example, “could have performed some other act,” or, according to a weaker reading, it could just mean “could have omitted performing the act in question.” On the face of it, it seems that an agent could have the ability to omit performing a certain act without having the ability to perform any other act, and it seems plausible to think that the ability to omit performing the act in question could be all that is required for responsibility for the act.⁴ This is, then, the interpretation of the principle of alternative possibilities that I will adopt here:

(PAP) A person is morally responsible for performing a given act only if he could have omitted performing that act.

According to PAP, the relevant alternatives or forking paths from which an agent acts when he acts freely are performing the act in question and omitting to perform it. One is the actual alternative

² In “Actuality and Responsibility” (*Mind*, cxx, 480 (October 2011): 1071–97), I argue that the metaphysics of causation plays an important role in the formulation of another popular view of responsibility (a main competitor to the view that is the focus of this paper): the “actual-sequence” view, according to which responsibility does not require alternative possibilities but is instead a function of actual causal sequences.

³ Harry Frankfurt, “Alternate Possibilities and Moral Responsibility,” this JOURNAL, LXVI, 23 (Dec. 4, 1969): 829–39.

⁴ See, for example, David Widerker, “Libertarianism and Frankfurt’s Attack on the Principle of Alternative Possibilities,” *The Philosophical Review*, CIV, 2 (April 1995): 247–61; and Helen Steward, “Fairness, Agency, and the Flicker of Freedom,” *Noûs*, XLIII, 1 (March 2009): 64–93.

(the alternative that the agent actually chooses), and the other is a counterfactual alternative (an alternative that the agent could have chosen but did not choose).

Van Inwagen suggested ways of extending the forking-paths metaphor to the other two main varieties of responsibility: responsibility for omissions and for outcomes in the world (events and states of affairs that are consequences of the agent's acts and omissions).⁵ In the case of omissions, the metaphor is cashed out in terms of the "Principle of Possible Action":

(PPA) A person is morally responsible for failing to perform a given act only if he could have performed that act.

And, in the case of outcomes, it is cashed out in terms of the "Principle of Possible Prevention":

(PPP) A person is morally responsible for an outcome (that is a consequence of one of his acts/omissions) only if he could have prevented that outcome.⁶

Now, is there a common idea captured by these three principles of responsibility? Is there a general picture of freedom and responsibility that they articulate? If so, what is it? These questions have not been examined carefully in the literature. Part of the problem is that, whereas a lot of thought has been put into the principle for acts (PAP), the other two principles have received much less attention (in particular, PPP is seldom the focus of any discussion), and, as a result, it has been hard to keep track of the general, unifying picture of freedom that the forking-paths metaphor allegedly depicts.

The plan for the paper is this. I will start by taking a closer look at that general picture. I will suggest that a natural way to identify a common thread in the three principles of responsibility is to understand all three principles in causal terms. Once the three principles are reformulated in these terms, it is possible to extract a general principle of freedom and responsibility. However, a better look at that general principle and some problems it faces will motivate an alternative way of understanding the forking-paths metaphor. Not only

⁵ Peter van Inwagen, *An Essay on Free Will* (New York: Oxford, 1983).

⁶ This formulation combines two separate principles offered by van Inwagen: a principle for events, PPP1, and a principle for states of affairs, PPP2. On events, see also John Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge, 1998), p. 99. (Fischer and Ravizza distinguish a stronger version, according to which responsibility for an event requires the ability to bring about some event incompatible with it, from a weaker version according to which it requires the ability to bring about the absence of the actual event. This is parallel to the distinction between the strong and weak versions of PAP. Again, the weaker version seems preferable.)

does causation play a central role in that alternative view too, but, as we will see, the alternative view also draws heavily on some metaphysical assumptions about the nature of causation. In fact, it is precisely in virtue of those assumptions that it promises to paint a more plausible picture of freedom and responsibility than the original view.

A main aim of the paper, then, is to motivate a new version of the alternative-possibilities view and to identify the metaphysical assumptions about causation on which it rests. My ultimate aim is not to argue for the truth of this view, however. In fact, towards the end of the paper I will suggest that, although the new model is a significant improvement over the traditional model, there is good reason to think that it is still deficient. I will suggest, in particular, that the motivation behind it can be captured in terms that are consistent with the rejection of that model.

II

What could the three principles of responsibility (PAP, PPA, and PPP) have in common?

Here is a suggestion. The principle about outcomes, PPP, appears to make reference to causation. “Preventing” an outcome is usually understood as causing or bringing about its absence. Read in that way, PPP claims that responsibility for an outcome requires having the ability to bring about its absence. So imagine that we understand acting and omitting to act along the lines of a causalist view of agency (a widespread view of agency which we will simply take for granted, in very broad terms, for the purposes of this paper). Imagine, for example, that we understand acts and omissions in the way Davidson does, as events (or absences of events) with a certain causal ancestry, in particular, one that includes certain mental events or states of the agent such as intentions, beliefs, desires, and so on.⁷ On this view, when an agent performs a certain act *A*, he brings about *A* (the relevant mental events or states bring about *A*), and, similarly, when an agent omits to perform a certain act *A*, he brings about the absence of *A* (the relevant mental events or states bring about the absence of *A*).

So consider PAP, the principle that says that responsibility for an act *A* requires the ability to omit to *A*. Given that on a causalist view omitting to *A* requires bringing about the absence of *A*, PAP claims that

⁷ See Donald Davidson, “Actions, Reasons, and Causes,” this JOURNAL, LX, 23 (Nov. 7, 1963): 685–700. In “Omissions and Causality” (*Noûs*, XLIII, 3 (September 2009): 513–30), I argued that the relevant causal ancestry of omissions is significantly different from that of actions, and that this makes trouble for attempts to capture omissions in traditional causalist terms. Still, a less orthodox causalist view could still be true of omissions.

responsibility for an act *A* requires the ability to bring about the absence of *A*. Similarly, take the principle for omissions, PPA (responsibility for omitting to perform a given act *A* requires the ability to perform *A*). Given that on a causalist view performing act *A* requires bringing about *A*, PPA claims that responsibility for omitting to *A* requires the ability to bring about *A*. But *A* is, in other words, the absence of the omission to *A*. Thus PPA claims that responsibility for omitting to *A* requires the ability to bring about the absence of the omission to *A*.

The causalist assumption,⁸ then, allows us to extract the following general principle of responsibility out of the three individual principles (“AP” is for “Alternative Possibilities”):

(AP) A person is morally responsible for *X* only if he could have brought about the absence of *X*.

AP provides a general picture of responsibility and a clean way of understanding the “forking paths” in the forking-paths metaphor. On this view, responsibility requires a kind of dual control that is represented by the existence of two alternatives that are open to the agent. These alternatives are: *making something happen* and *making it not happen*.

Now, the alternative-possibilities model has been subject to pressing criticisms. In the next section I examine some of the alleged counterexamples; I explain how they threaten to undermine AP; and I use them to motivate a competing principle of alternative possibilities.

III

I will start by looking at the case of responsibility for outcomes. As noted before, van Inwagen’s principle for outcomes, PPP, has received very little attention in the literature. My reason for focusing on outcomes first is that, as we will see, it is easier to motivate the new principle of alternative possibilities in this case than in the case of acts and omissions (although it is not hard to see how the principle applies to acts and omissions once it is in place).

According to AP, agents are only responsible for outcomes that they could have prevented. But there seem to be obvious counterexamples to this idea. Consider, first, a scenario of (symmetric) overdetermination:

TWO ASSASSINS: Two assassins maliciously shoot at a victim at the same time. The victim dies. Each bullet would have been independently sufficient for the death. The assassins were not aware of each other’s plans, and they could not have stopped each other.

⁸Note that I am including as part of the causalist assumption the claim that absences can enter in causal relations. This assumption played a role above too, when we understood preventing an outcome as causing its absence.

Here the death is overdetermined by the two assassins' acts. Arguably, each assassin is morally responsible for the victim's death in this case. However, neither assassin could have prevented the death, given the presence of the other assassin. So TWO ASSASSINS seems to be a counterexample to AP.

Now, here is a natural thing to say about this case: although the assassins could not have prevented the death, there is something that they could have done: they could have at least *failed to bring it about*. If Assassin 1, say, had failed to shoot, then the victim would have died in the hands of Assassin 2, but then Assassin 2, not Assassin 1, would have caused the victim's death.⁹ This very natural thought motivates a revision of the principle of alternative possibilities. Perhaps responsibility for *X* does not require the ability to bring about *X*'s absence but only the ability to fail to bring about *X*. Perhaps the two "forking paths" that must be available to the agent in order for him to have the relevant kind of dual control are not making something happen and *making it not happen* but, instead, making something happen and *not making it happen*.

Let us call this new principle "AP*." On a first pass, the principle says:

(AP*-First Pass) A person is morally responsible for *X* only if he could have failed to bring about *X*.

AP* and AP differ in what they take the relevant counterfactual alternative to be: whereas AP takes it to be the prevention of *X*, AP* takes it to be the failure to bring about *X*'s occurrence. AP* requires less from the counterfactual alternative than AP. For preventing *X* entails failing to cause *X*, but failing to cause *X* does not entail preventing *X* (*X* could still occur, for other reasons¹⁰).

Not only is AP* consistent with our intuitive judgment about responsibility in overdetermination cases, but it also seems to capture an important insight about responsibility. Roughly, this is the thought that responsibility for something requires the ability to *fail to be involved* in the occurrence of that thing. For example, it is natural to think that

⁹The claim is not that, by failing to shoot, the assassin would have failed to bring about the victim's death because failing to shoot is an omission. Recall that I am working under the assumption that omissions can be causes (see note 8). Even if omissions can in principle be causes, this is not a scenario where an omission would be a cause.

¹⁰On a certain view of events (embraced, for example, by van Inwagen in *An Essay on Free Will*), events have their causes essentially. Naturally, if this view were true, an event that is actually caused by an agent could not occur for any other reasons. I will not argue against this view here, since I agree with David Lewis that it has very little initial plausibility (see Lewis, "Causation," in *Philosophical Papers, Volume II* (New York: Oxford, 1986), pp. 159–213; see especially n. 20).

the two assassins are responsible for the outcome because, although they chose to cause the outcome, or to be involved in the occurrence of the outcome, they could have chosen not to cause it, or to fail to be involved in its occurrence. Clearly, this is an ability that agents can have regardless of whether the outcome would still have occurred if they had not brought it about. What determines an agent's responsibility for something is, intuitively, not what actually happens, but what the agent's relation to what actually happens is. So an agent can fail to be involved in the occurrence of an outcome, and can fail to be responsible for it, even if the outcome still occurs. This is the idea that AP* strives to capture. AP, on the other hand, requires more for the existence of responsibility: it requires, in addition to the ability to fail to be involved in the bringing about of *X*, the ability to *be* involved in the bringing about of *X*'s absence. On reflection, this seems unmotivated.

In section 1, we distinguished between a stronger reading and a weaker reading of "the ability to do otherwise." Basically, the distinction was that between the ability to bring about something other than *A* (say, *B*) and the ability to bring about the absence of *A*. We pointed out that, on the face of it, one can have the ability to bring about the absence of *A* without having the ability to bring about some other act *B*, and that it is probably too strong and unmotivated to claim that responsibility requires the stronger ability. We can now see that even the weaker reading is probably too strong. For, just like there is a distinction between the ability to bring about *B* and the ability to bring about the absence of *A*, there is a similar distinction between the ability to bring about the absence of *A* and the ability to fail to bring about *A*. On the face of it, one can have the ability to fail to bring about *A* without having the ability to bring about the absence of *A*. And, again, it is probably too strong and unmotivated to claim that responsibility requires the stronger ability.

Let us look at some more examples. Consider now a scenario of "preemption," along the lines of one offered by van Inwagen:¹¹

GUNNAR AND RIDLEY: Gunnar maliciously shoots and kills Ridley. Unbeknownst to Gunnar, a backup assassin, who also wanted Ridley to die, would have shot Ridley if Gunnar had not done so himself. Gunnar would not have been able to stop the backup assassin.

¹¹ Van Inwagen, *op. cit.*, p. 170. Van Inwagen only gives the example in order to respond to it, since he believes in PPP's truth. The example is intended to mimic the structure of "Frankfurt-style cases," the scenarios that are usually wielded against PAP (we will consider these in the next section). Van Inwagen offers two different versions of the example. I will examine the second version separately, later in this section.

Here, too, there are two possible routes to the death, but this time only one of them is active. The active causal route “preempts” the alternative route. GUNNAR AND RIDLEY seems to be a counterexample to AP too, for, intuitively, Gunnar is responsible for Ridley’s death although he could not have prevented it. He could not have prevented it because of the backup assassin’s existence; however, he is still responsible for it because the backup assassin never had to intervene.

Van Inwagen has an ingenious response, which I will not get into here. For, again, even if GUNNAR AND RIDLEY is a counterexample to AP, it is not a counterexample to AP*. Even if Gunnar could not have prevented Ridley’s death, he could have failed to cause it. If the backup assassin had to intervene, then the death would still have occurred, but in that case the backup assassin, not Gunnar, would have caused it.

At this point the following objection might be raised: “But, if Gunnar’s failure to shoot would have triggered the backup assassin’s shooting, then, given that the backup assassin’s shooting would have in turn resulted in Ridley’s death, it follows by the transitivity of causation that Gunnar’s failure to shoot would have caused the death. So Gunnar could not have failed to cause the death: he would still have caused it by failing to shoot, albeit indirectly.”¹²

However, the transitivity of causation has been challenged in recent years, and by appeal to scenarios that have precisely the same kind of causal structure. Consider, for example:

BOULDER: A boulder is dislodged and starts rolling towards Hiker. Hiker notices the boulder and ducks in response. Hiker survives.¹³

This is a case where the transitivity of causation allegedly fails: the boulder’s rolling towards Hiker causes Hiker to duck, which causes Hiker to survive, but, presumably, the boulder’s rolling towards Hiker does not cause Hiker’s survival. The reason we feel that the boulder’s rolling towards Hiker does not cause his survival is, intuitively, that it creates a “threat” to his survival, one that has to be

¹² Another objection that might be raised at this point is: “Couldn’t we change the example so that the backup assassin would have intervened by making Gunnar a link in the causal chain leading to the death?” I consider this possibility later in this section.

¹³ This example, which is discussed in Christopher Hitchcock’s “The Intransitivity of Causation Revealed in Equations and Graphs” (this JOURNAL, xcvi, 6 (June 2001): 273–99), is originally due to Ned Hall (from an early draft of his “Two Concepts of Causation,” in John Collins, Hall, and L. A. Paul, eds., *Causation and Counterfactuals* (Cambridge: MIT, 2004), pp. 225–76).

countered in order for him to still survive. The creation of the threat causes the cancellation of the threat, and the cancellation of the threat causes the final outcome, but, presumably, the creation of the threat does not itself cause the outcome. (This is so even if the same thing that launches the threat helps to cancel it. Intuitively, if something cancels a threat that it itself created, it does not make a causal contribution to the outcome that ensues.)

Let us call scenarios of this kind *threat-cancellation scenarios*.¹⁴ It is hard to spell out exactly what the relevant concept of threat is in threat-cancellation scenarios. But one thing that seems clear is that it is not an indeterministic concept; that is, the boulder is not a threat in the sense that it makes Hiker's survival less objectively likely. For we would still think that the boulder is not a cause of Hiker's survival, and for what is basically the same reason, if determinism were true and it was determined that Hiker would survive. The relevant sense of threat is one according to which the rolling boulder is a threat to Hiker's survival (one that needs to be countered, and one that will be countered, for Hiker to survive) in a fully deterministic world. For the purposes of this paper, we will have to rely on intuition to pick out the relevant notion of threat.¹⁵

Now, GUNNAR AND RIDLEY seems to have a threat-cancellation structure too. Here the outcome is Ridley's death. If Gunnar decides not to shoot, this creates a threat to the occurrence of that outcome, and, as a result, the threat has to be countered (by the backup assassin) so that the death still occurs. Again, the creation of the threat causes its cancellation, and the cancellation of the threat in turn causes the death, but, presumably, the creation of the threat does not cause the death. So Gunnar does not cause Ridley's death if he decides not to shoot Ridley. And, again, the relevant concept of threat is one that is consistent with determinism: Gunnar's failure to shoot intuitively creates a threat to Ridley's death, and thus fails to cause Ridley's death, even if it does not make the death any less

¹⁴Hall calls these structures "short circuits" (see "Causation and the Price of Transitivity," this JOURNAL, xcvi, 4 (April 2000): 198–222, section II; and "Structural Equations and Causation," *Philosophical Studies*, cxxxii, 1 (January 2007): 109–36, section 3.3). Stephen Yablo calls them "the metaphysical equivalent of Stockholm Syndrome" (see "Advertisement for a Sketch of an Outline of a Prototheory of Causation," in Collins, Hall, and Paul, eds., *op. cit.*, p. 123). (Stockholm Syndrome is the gratitude that victims of kidnapping sometimes feel towards their kidnappers when the kidnappers help them with problems caused by the captivity, that is, when they help to neutralize some threats that they themselves created.)

¹⁵Some of the works cited in note 17 contain attempts to elucidate that notion.

objectively likely—that is, even if it was determined that Ridley would still die.¹⁶

Of course, the transitivity or intransitivity of causation is a contested issue, and I do not intend to settle that debate here. All I intend to do by bringing this up is draw attention to the fact that the advocate of the new model of alternative possibilities has readily available to him an argument of this kind. It is a line of argument that has much initial plausibility and is supported by some recent work in the metaphysics of causation.¹⁷

TWO ASSASSINS and GUNNAR AND RIDLEY are just two examples where AP and AP* yield different results. But there are many other examples like them, all involving backup causes of some kind. Consider a scenario taken from another context: Bernard Williams's famous "George the Chemist" case, which he offers as a counterexample to utilitarianism.¹⁸ George is an unemployed chemist looking for work, and the only job he can find is in the production of chemical weapons, which he knows would result in significant harm to humanity. He knows that, if he does not take the job, someone else will, and the harm will still be done. So George cannot prevent the harm from being done. Still, imagine that he decides to take the job. In that case we will likely hold him responsible (blameworthy) for the ensuing harm. By appeal to AP*, an advocate of the alternative-possibilities view can agree that George is blameworthy for the harm. For he chose to cause the harm when he could have failed to cause it. If he had not taken the job, then someone else (whoever is hired in his place) would have caused the harm, and he would not have. Again, the claim would be that George would not have caused the harm by not taking the job even if his not taking the job would have resulted in someone else's taking the job, and even if this, in turn, would have resulted in the harm.

¹⁶This is relevant because the question that underlies this debate is, again, whether the truth of determinism is consistent with the kind of freedom required by moral responsibility. Many advocates of the alternative-possibilities model of responsibility argue that it is not. If so, their arguments presumably cannot presuppose the failure of determinism.

¹⁷For discussion of the transitivity of causation, see, for example, Hall, "Causation and the Price of Transitivity," "Two Concepts of Causation," and "Structural Equations and Causation"; Hitchcock, *op. cit.*; Carolina Sartorio, "Causes as Difference-Makers," *Philosophical Studies*, cxxiii, 1/2 (March 2005): 71–96, and "On Causing Something to Happen in a Certain Way without Causing It to Happen," *Philosophical Studies*, cxxix, 1 (May 2006): 119–36; Yablo, "De Facto Dependence," this JOURNAL, xcix, 3 (March 2002): 130–48, and "Advertisement for a Sketch of an Outline of a Prototheory of Causation."

¹⁸Bernard Williams, "A Critique of Utilitarianism," in J. J. C. Smart and Williams, *Utilitarianism: For and Against* (Cambridge, UK: University Press, 1973).

For his not taking the job would have created a threat to the occurrence of harm which had to be cancelled in order for the harm to still occur. It is plausible to believe that this is, in essence, the intuition that Williams was trying to capture with the example: George should not take the job, and will be blameworthy if he does, because, even if the harm will ensue regardless of what he does, he can still decide if he wants to be *involved in* the production of that harm. Again, this is the intuition that AP* aims to capture.

As a final example, consider a scenario used by John Fischer to illustrate the difference between “regulative control” (the kind of control that he takes to be captured by the forking-paths metaphor) and “guidance control” (the kind of control that he thinks is actually required for responsibility). The driver of a car decides to steer the car to the right, turns the steering wheel accordingly, and causes the car to go right in the normal way. According to Fischer, the driver had regulative control of the car’s turning right only if he had the power to steer the car in a different direction. If it turns out that the mechanisms to steer the car in a different direction are broken, the driver fails to have the kind of dual control required by regulative control (although he can still have guidance control of the car, to the extent that the mechanism to turn right is working properly).¹⁹ Similarly in the case of a student driver where, if the student had tried to steer the car in a different direction, the driving instructor would have intervened and caused the car to go right: the student does not have regulative control of the car’s turning right when she makes it go right, for she could not have steered the car in any other direction.²⁰ Again, by appeal to AP*, an advocate of the alternative-possibilities model can grant that the agent is responsible for the outcome. For, although she could not have made the car go in a different direction, she could have failed to make the car go right. If she had tried to make the car go in a different direction (by turning the steering wheel accordingly), then the car would still have gone right, but she would not have caused the car to go right. The failure of the relevant mechanism would have, or the driving instructor would have, but she herself would not have, for she would have created a threat that had to be countered for the car to still go right.²¹

¹⁹ See Fischer, “Responsibility and Alternative Possibilities,” in Widerker and Michael McKenna, eds., *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (Burlington, VT: Ashgate, 2003), pp. 27–52, at p. 28.

²⁰ See *ibid.*, p. 49n3.

²¹ Yet another kind of example where AP* would let us say that the agent is responsible and AP would not is one involving outcomes that are collectively caused but

So AP* draws a different picture of the alternative possibilities required by responsibility, and one that is arguably a more promising picture than that offered by AP. The metaphysics of causation plays a key role in completing that picture: the intransitivity of causation, in particular, helps make sense of the idea that agents can *fail to be involved* in the occurrence of certain outcomes even when they trigger certain backup mechanisms that result in those outcomes.

Now, as it stands, AP* only works as a first pass. For, despite being significantly weaker than AP, it is still too strong. To see this, imagine a variant on GUNNAR AND RIDLEY. As I was imagining the original scenario, the backup assassin would have shot Ridley *himself* if Gunnar had not shot him (thus he would have caused Ridley's death, and Gunnar would not have). But imagine that the backup assassin would have forcibly made *Gunnar* shoot Ridley if Gunnar had decided not to do so (say, the backup assassin is a big guy, so he could have easily forced Gunnar's fingers to squeeze the trigger in the required way).²² If the causal chain goes through Gunnar in this way, then Gunnar is undoubtedly one of the causes of the death. But then it seems that Gunnar could not have failed to cause Ridley's death in this case: if he had chosen not to shoot Ridley, he would still have shot him, manipulated by the backup assassin. Hence, it follows from AP* on its first pass that Gunnar is not responsible for Ridley's death in this case. However, it seems wrong to suggest that Gunnar is any less in control of, or that he is any less responsible for, Ridley's death when the backup assassin would have intervened by forcing him to shoot Ridley than when he would have intervened by shooting him himself. How could this possibly make a difference to his responsibility?

But there is a natural response to this. It is to point out that, although Gunnar could not have failed to cause Ridley's death in this variant

also partially overdetermined. Consider the effect that driving certain types of cars allegedly has on the environment (for example, pollution or global warming). It is plausible to think that I cannot prevent that outcome by failing to drive my environmentally unfriendly car, since so many other people are still going to drive their own environmentally unfriendly cars and the same outcome will result. Still, I can fail to be one of the contributing causes. (Thanks to Cheshire Calhoun for the example.) Consider, also, an agent's casting a vote in an election where the candidate whom the agent voted for wins: the agent could have failed to contribute to that candidate's winning by failing to vote for him, although (in all realistic cases) he could not have made a difference to the outcome of the election.

²²This is the second version of the example suggested by van Inwagen (see n. 11). Another type of example that makes trouble for the principle as it stands is a Frankfurt-style case where the backup mechanism would have intervened even earlier, at the level of choices. I discuss a scenario of this kind in the next section.

of the case, there is something in the causal chain leading to the death (something that he also caused and in virtue of which he is responsible for the death) that he *could* have failed to cause, namely, his shooting of Ridley. If he had chosen not to shoot Ridley, he would have ended up shooting him all the same, but he would not have been a cause of the shooting. For he would have created a threat to the shooting that needed to be countered (by the backup assassin) for it to still occur. In other words, the new variant of the case is also a threat-cancellation scenario: it is a threat-cancellation scenario with respect to, not the event of Ridley's death, but Gunnar's earlier shooting of Ridley.

This suggests the following revision of the principle:

(AP*) A person is morally responsible for *X* only if he could have failed to bring about *X* or some cause of *X* that he also caused and in virtue of which he is responsible for *X*.

This revision seems well motivated: if I could have failed to cause some event in the chain leading to an outcome and I still did not, it seems that I can have the relevant kind of dual control that is allegedly necessary for responsibility, even if I could not have failed to cause the outcome itself. For, appealing to the forking-paths metaphor again, there is still a point in the road where the road forks into two different paths, and where I could have taken one path rather than the other. Similarly, when Gunnar, in the new version of the case, chooses not to shoot Ridley, he still chooses to fail to be involved in the process leading to Ridley's death, even if he cannot choose to fail to be involved in the bringing about of the death itself. So, an advocate of the alternative-possibilities view could argue, he still has the relevant kind of dual control that is required for responsibility.²³

IV

Now that the new principle of alternative possibilities AP* is in place, let us see how it applies to acts and omissions.

In the case of acts AP* says that a person is responsible for an act, which according to a causalist view of agency is an event of a certain

²³ Advocates of the traditional model of alternative possibilities sometimes include a "tracing condition" in their principles of responsibility too, in order to avoid similarly unwelcome results (see, for example, David P. Hunt, "Moral Responsibility and Buffered Alternatives," *Midwest Studies in Philosophy*, xxix, 1 (September 2005): 126–45, who does this in the case of the principle for actions). Others prefer to restrict their principles to cases of "nonderivative" responsibility (cases where the agent is responsible for his act but not in virtue of being responsible for some earlier act).

kind brought about by the agent (by means of the relevant mental events or states), only if he could have failed to bring about that event or some cause of it that he brought about and in virtue of which he is responsible for the act. This is by contrast with AP, which claims that responsibility for an act requires having the ability to *prevent* the relevant event.

The traditional principle for acts has been challenged by the famous “Frankfurt-style scenarios,” where an agent is arguably responsible for an act that he could not have failed to perform.²⁴ Here is a classical example:

JONES AND SMITH: Jones deliberates about whether to shoot Smith and arrives at the decision to shoot him completely on his own. He kills Smith. An evil neuroscientist had been monitoring Jones’s brain closely. Had Jones been moved by the reasons not to shoot Smith, the scientist would have intervened by manipulating Jones’s brain directly, and, as a result, Jones would still have chosen to shoot Smith.

As Frankfurt pointed out, it seems that in this case Jones could not have failed to shoot Smith (given the scientist’s presence), but he is still responsible for his act of shooting Smith, given that he decided on his own to shoot him (the scientist never had to intervene).

Now, is JONES AND SMITH a problem for AP* as well? This depends on whether Jones could have failed to bring about the event consisting in his shooting Smith or some relevant link in the chain leading to that event.

First, then, could Jones have failed to bring about his shooting Smith? It seems that he could not have. For, in the counterfactual scenario where the scientist intervenes, Jones’s body moves in the same way and as a result of Jones’s choosing to pull the trigger (this is so even if the choice itself is manipulated by the scientist).²⁵ But consider, now, the causes of the bodily movement that he actually brought about, and in virtue of which he is responsible for his act. Consider, in particular, his choosing to shoot Smith. Could Jones have failed to bring about this link in the causal chain (which he actually brought about by engaging in a process of deliberation that resulted in the choice)? It seems that, indeed, he could have. If he had given any serious consideration to the reasons not to kill Smith, then the scientist would have intervened by manipulating his brain

²⁴ See Frankfurt, *op. cit.*

²⁵ If the scientist would have intervened by directly manipulating Jones’s *body* instead of his choices, then, I would argue, Jones would not have brought about his own shooting of Smith.

in such a way that he would have ended up making the choice anyway.²⁶ Still, arguably, in that case he would not have *caused* the event consisting in his making that choice (only the scientist would have). For, once again, he would have created a threat to the occurrence of that event that needed to be countered, by the scientist, for it to still occur. Arguably, then, JONES AND SMITH is not a counterexample to AP*.

The same goes for similar examples where the scientist would have intervened earlier in the chain leading to Jones's choice. Imagine, for example, that the scientist would have intervened by implanting in Jones's brain the judgment that the reasons to kill Smith override the reasons not to kill him. Then the scientist would have brought about that judgment by Jones; Jones himself would not have. In general, take whatever link in the chain is guaranteed by the presence of the scientist. It seems that, although Jones could not have prevented the occurrence of that event, it will always be the case that, if the scientist had intervened, Jones would not have brought about that event.²⁷

A nice feature of AP*, then, is that it appears to be immune to the standard challenge posed by Frankfurt-style cases. Something similar holds for omissions. Many people seem to think that Frankfurt-style cases undermine the alternative-possibilities requirement for responsibility in the case of omissions too.²⁸ But, again, AP* seems to be immune to the challenge. AP* says that a person is morally responsible for an omission, which on a causalist view is the absence of an event of a certain kind brought about by the agent (by means of the relevant mental events or states), only if he could have failed to bring about that absence or some cause of it that he brought

²⁶ I am assuming that the result of the manipulation would have been a genuine choice by Jones and that it would have been the same token event as the one that actually occurred when he made the choice on his own. Of course, if this were not the case, then Jones could have failed to bring about that event for the simple reason that it would not even have occurred if the scientist had intervened.

²⁷ It might be objected that, if the scientist would have intervened early enough in the chain, upon being triggered by an event *over which Jones had no voluntary control*, then the fact that Jones would not have caused the ensuing link in the chain is not enough to ground his responsibility. I discuss this objection in the next section.

²⁸ See, for example, Frankfurt, "An Alleged Asymmetry between Actions and Omissions," *Ethics*, CIV, 3 (April 1994): 620–23; Randolph Clarke, "Ability and Responsibility for Omissions," *Philosophical Studies*, LXXIII, 2/3 (March 1994): 195–208; Alison McIntyre, "Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency," *Philosophical Review*, CIII, 3 (July 1994): 453–88; and Fischer and Ravizza, *op. cit.*, chapter 5. In this case I think it is less clear that Frankfurt-style cases are effective against AP (see Sartorio, "A New Asymmetry between Actions and Omissions," *Noûs*, XXXIX, 3 (September 2005): 460–82).

about and in virtue of which he is responsible for the omission. So consider, as an example of a Frankfurt-style omission case, the flipside of the earlier Frankfurt-style action scenario. This time Jones decides on his own not to shoot Smith. A scientist (of the benevolent kind) has been closely monitoring his brain. Had Jones shown any signs of being about to choose otherwise, the scientist would have intervened by manipulating Jones's brain so that he chooses not to shoot Smith. Frankfurt would say: Jones is responsible (praiseworthy, perhaps) for not shooting Smith, but, given the presence of the scientist, he could not have shot him. But this is not a problem for AP*. For, again, a proponent of AP* would say that Jones could have failed to bring about a link in the causal chain issuing in his omission, say, his choice not to shoot Smith. He could have failed to bring about the relevant link in the causal chain because, had the scientist intervened, then the scientist only (not him) would have brought it about.

v

What emerges from the discussion of Frankfurt-style cases—and, in general, cases involving fail-safe mechanisms of various sorts—is this. By relying on AP*, an advocate of the alternative-possibilities view can argue that the agent has dual control in those cases. He has dual control because, although he could not have done otherwise, or decided to do otherwise, or prevented the events that followed, he could have failed to bring about the relevant events or states, such as the relevant outcomes in the world, or the relevant bodily movements, or the causes of those. Responsibility does not require the ability to *do* otherwise but, so to speak, the ability to *bring about* otherwise, or the ability to have a different causal impact on the world.

This view is importantly different from all the existing defenses of the alternative-possibilities view from the Frankfurt-style counterexamples. Let me briefly compare it with other responses that have been defended in the literature.

Some philosophers have argued that, although responsibility does not require the ability to do otherwise, it may require something like the ability to “avoid responsibility.”²⁹ According to views of this kind, the agent in a Frankfurt-style case is responsible for his act because he could have avoided responsibility for his act, although

²⁹ See McKenna, “Alternative Possibilities and the Failure of the Counterexample Strategy,” *Journal of Social Philosophy*, xxviii, 3 (December 1997): 71–85; Keith D. Wyma, “Moral Responsibility and Leeway for Action,” *American Philosophical Quarterly*, xxxiv, 1 (January 1997): 57–70; and Michael Otsuka, “Incompatibilism and the Avoidability of Blame,” *Ethics*, cviii, 4 (July 1998): 685–701.

he could not have avoided acting as he did (he could have avoided responsibility for his act because he would have failed to be responsible for his act in the scenario where the scientist intervenes). AP* is not strictly speaking inconsistent with views of this kind, but it is still preferable to them. For, if AP* were true, it would provide a simple and elegant explanation of *why* responsibility might require the ability to avoid responsibility. According to AP*, if responsibility requires the ability to avoid responsibility, it is only because it requires, more fundamentally, the ability to fail to be involved causally. This would arguably be a deeper and more satisfying explanation, given the widespread belief that responsibility is in some important way *grounded in* causation.

AP* is also preferable to other views that are able to accommodate Frankfurt-style cases but at what looks like an implausible metaphysical cost, for example, views according to which the act performed by the agent if the scientist had intervened would not have been the same token act (or, perhaps, an act at all).³⁰ These views presuppose highly controversial criteria for event or act individuation. The same goes for “agent-causal” views according to which responsibility does not require the ability to perform a different act or to make a different choice, but the ability to fail to be the agent-cause of one’s act or choice.³¹ According to these views, agents in Frankfurt-style cases have the requisite kind of dual control because, even if the scientist can make the agent perform the relevant act, or make the relevant decision, he cannot make the agent be the “ultimate source” (or the “agent-cause”) of his act or decision. In a certain respect, this is the type of view that comes closest to AP*, given that they both rest on the claim that the agent *fails to bring about* his choice in a Frankfurt-style scenario where the scientist intervenes.

³⁰ See Maria Alvarez, “Actions, Thought-Experiments and the ‘Principle of Alternate Possibilities’,” *Australasian Journal of Philosophy*, LXXXVII, 1 (2009): 61–81; Steward, *op. cit.*; and McKenna, “Alternative Possibilities and the Failure of the Counterexample Strategy.” There is a point where McKenna seems to be endorsing a view similar to mine. He imagines a Frankfurt-style case with him as an agent where the relevant act is his jumping into the air and claims, about that case: “I may be responsible for jumping into the air, but this is because I could have avoided the situation in which I bring it about that I jump into the air. And this is a genuine alternative that was available to me even if I could not bring it about that I do some other type of thing than jump into the air” (p. 74). However, he then goes on to explain that the reason he thinks he could have avoided the situation in which he brings it about that he jumps into the air is that, if the scientist had intervened, his act of jumping into the air would have been a different token act (pp. 76–77).

³¹ See, for example, Timothy O’Connor, *Persons and Causes: The Metaphysics of Free Will* (New York: Oxford, 2000); and William L. Rowe, “Alternate Possibilities and Reid’s Theory of Agent-causation,” in Widerker and McKenna, eds., *op. cit.*, pp. 219–34.

But the views are importantly different in that AP* does not rest on the problematic notion of agent-causation and the commitments usually associated with it. In particular, it does not rely on the claim that the intervention of an external manipulator never results in the agent's being a cause of his own choice (in the relevant sense of "cause"). Instead, it relies on the much more mundane suggestion that the ordinary relation of causation can fail to be transitive.³²

VI

I have argued that, by looking more closely into the nature of causation, it is possible to motivate a new model of alternative possibilities. The new model is a significant improvement over the traditional model in that it captures an important insight about responsibility that the traditional model is unable to capture: the thought that responsibility requires the ability to "fail to be involved." The new model captures this insight in purely causal terms, and, on the basis of some plausible metaphysical assumptions about the nature of causation, it avoids the standard challenge posed by Frankfurt-style cases (and similar scenarios involving overdetermination and preemption), which threatens to undermine the traditional model.

Now let us assume that I am right and that AP* is indeed the best version of the alternative-possibilities model. If so, the fate of the alternative-possibilities model is tied to AP*'s fate. But is AP* true? A full answer to this question would deserve a much more extended discussion than the one I can offer here. Still, in what follows I briefly explain some reasons for being skeptical, and I offer some tentative conclusions.

A first reason to be skeptical about AP* is that, even if the *standard* Frankfurt-style objection fails to undermine AP*, a different objection based on Frankfurt-style cases (typically known as the "robustness objection") might still succeed. Fischer, among others, argued that the existence of alternative possibilities could only ground responsibility if the relevant alternatives were sufficiently "robust."³³ According to Derk Pereboom, this means, at least, that the alternative possibilities

³² Agent-causal chains are by their very nature chains that can only go through the agent himself. Event-causal chains are not like this. For example, I can easily event-cause my choice to miss work on a given day by causing other people to persuade me to go on a daytrip with them, even if this is done by means of brainwashing or other forms of manipulation. The feature of Frankfurt-style cases in virtue of which it is plausible to think that, if the scientist has to intervene, the agent does not cause the relevant event in the chain is not the fact that the chain involves another agent but the fact that it is a threat-cancellation scenario (the scientist cancels the threat that the agent himself created).

³³ See Fischer, "Responsibility and Alternative Possibilities."

that were open to the agent were such that the agent understood that he would avoid responsibility for his act by securing one of those alternative possibilities.³⁴ If the agent lacked such an understanding, the availability of those alternatives would be irrelevant to the agent's responsibility in the actual scenario. If this is right, then it means that AP* does not capture the whole content of the alternative-possibilities requirement on responsibility. The full statement of that requirement would have to be the claim that responsibility requires, not only the ability to do something in virtue of which you would fail to bring about, say, your choice, but, also, an understanding that you would be avoiding responsibility for your choice by doing that very thing.

It could be argued that, in a standard Frankfurt-style case like JONES AND SMITH, the agent does indeed have such robust alternative possibilities. For Jones could have at least started the process to decide not to shoot Smith, in which case he would have failed to bring about his choice to shoot Smith, *and* he understood (or it is reasonable to think that someone in his position could understand) that by starting that process he would avoid responsibility for shooting Smith. But it is possible to imagine other Frankfurt-style scenarios where, arguably, no such robust alternative possibilities are open to the agent. Consider, for example, the following variation on JONES AND SMITH, which closely resembles a scenario discussed by Pereboom:³⁵

JONES AND SMITH*: The scientist knows Jones's psychological profile well. He knows that, for Jones to choose *not* to shoot Smith, it is causally necessary that a moral reason not to shoot him occur to him with a certain force (either involuntarily or as a result of his voluntary activity). Even though Jones could not choose not to shoot Smith without that moral reason occurring to him, were such a moral reason to occur to him at any point, as far as Jones is aware he could still choose to shoot Smith by deciding to ignore that reason. However, the scientist has implanted his device in Jones's brain to ensure that Jones decides to shoot Smith. The device would be triggered only if it sensed that the relevant moral reason not to shoot Smith occurred to Jones with the specified force. As a matter of fact, that moral reason never occurs to Jones, so the scientist never has to intervene.³⁶

Again, intuitively, Jones is responsible for shooting Smith because the scientist did not have to intervene. Given that a moral reason

³⁴ See Derk Pereboom, *Living without Free Will* (New York: Cambridge, 2001), chapter 1.

³⁵ *Ibid.*, p. 22.

³⁶ Not only is this supposed to be a scenario of responsibility without robust alternative possibilities, but it is also designed to avoid the famous "dilemma objection"

not to shoot Smith could have occurred to Jones, perhaps even as a result of his own voluntary activity, Jones *could* have failed to bring about his choice to shoot Smith in this case (for, as we have seen, if the scientist had intervened, Jones would not have brought about his choice as a result). Still, Pereboom (and others) would argue, this alternative possibility is not sufficiently robust (even if Jones himself could have voluntarily made the moral reason not to shoot Smith occur to him with the specified force). For, given that Jones was not aware of the scientist's existence, as far as he could tell he would not have avoided responsibility for his act of shooting Smith simply by making that moral reason occur to him. This is so because we are assuming that, as far as he could tell, he could still have decided to ignore that reason and choose to shoot Smith. Hence, the objection would go, AP* fails. Responsibility is not grounded in any important way in the existence of alternative possibilities; in particular, it is not grounded in the ability to fail to bring about the relevant choice.

This is not the place to assess the force of this objection.³⁷ All I want to note is that this is an objection to the alternative-possibilities view that is not addressed by the new model. Hence, even if AP* is preferable to AP, AP* may still fail. In fact, if I am right and AP* is the best interpretation of the forking-paths metaphor, and if the robustness objection still manages to undermine it, then this is bad news for the alternative-possibilities view of responsibility.

A second reason to be skeptical about AP* is that it is possible to argue that the main guiding intuition behind AP*, the thought that responsibility requires "the ability to *fail to be involved*," is something that can be captured, at least partly, within a view according to which responsibility does not require any alternative possibilities (or any robust alternative possibilities). Take JONES AND SMITH*, the scenario that allegedly shows that responsibility does not require any robust alternative possibilities. We have seen that, in this case: (a) Jones seems to be responsible for his act of shooting Smith, (b) he had the ability to fail to bring about his choice to shoot Smith, but (c) the ability

to Frankfurt-style cases developed by Widerker in "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities." See also the scenarios discussed by Mele and David Robb in "Rescuing Frankfurt-Style Cases" (*Philosophical Review*, cvii, 1 (January 1998): 97–112); and by McKenna in "Robustness, Control, and the Demand for Morally Significant Alternatives: Frankfurt Examples with Oodles and Oodles of Alternatives" (in Widerker and McKenna, eds., *op. cit.*, pp. 201–17).

³⁷ For further discussion of the robustness objection, see Fischer, "Recent Work on Moral Responsibility," *Ethics*, cx, 1 (October 1999): 93–139, section III; and Pereboom, *op. cit.*, chapter 1.

to fail to bring about his choice does not seem to *ground* his responsibility for his choice, or for his act. I think that there is a sense in which, even if the failure to bring about his choice does not ground his responsibility for his choice (or for his act) in the way that AP* would require, it *does* ground it in a way that is consistent with the *rejection* of AP*. Moreover, this fact could explain, at least partly, why the alternative-possibilities model might seem attractive while perhaps being a deficient model of responsibility.

Let me explain. In JONES AND SMITH*, Jones is actually responsible for his choice of shooting Smith. Given that considering the moral reasons not to shoot Smith was, in the circumstances, necessary for him not to make the choice to shoot Smith, it seems that Jones is responsible for his choice to shoot him, at least in part, in virtue of *not* having considered those reasons. Since the scientist would have intervened if he had considered those reasons, Jones would not have been responsible for his choice if he had considered them. Now, as we have seen, considering those reasons would not by itself have absolved Jones of responsibility for choosing to shoot him, since, as far as he could tell, he could still ignore those reasons. But it still seems right to suggest that, to the extent that he is responsible for choosing to shoot Smith in virtue of having *failed* to consider those reasons, he could not have been responsible for choosing to shoot him in virtue of having *considered* them. More precisely, it seems to me that considering those reasons would not have made Jones responsible for the choice to shoot Smith, because it would not have contributed to *bringing about* that choice. In other words, part of what grounds the fact that Jones is responsible for his choice to shoot Smith (in virtue of having failed to consider moral reasons not to shoot him) is that, if he had considered those reasons, doing so would not have contributed to bringing about his choice to shoot Smith. In the closest possible world(s) where he considers those reasons, considering those reasons does not bring about his choice to shoot Smith. Moreover, the fact that, in the closest possible world(s) where he considers those reasons, considering those reasons does not bring about his choice to shoot Smith, helps ground Jones's responsibility for his choice in the actual world.

We can see this point more clearly by appeal to the GUNNAR AND RIDLEY scenario described earlier in the paper. As we have seen, according to AP*, Gunnar is responsible for Ridley's death because, although he could not have prevented his death, he could have failed to bring it about: the fact that he had the ability to bring about the death is part of what makes him responsible. But, again, it could be argued that what grounds his responsibility is not this but something

in its vicinity. Namely, it is the fact that *if he had failed to shoot Ridley, his failure to shoot him would have failed to bring about his death*. In other words, in the closest possible world(s) where Gunnar does not shoot Ridley, Gunnar fails to cause Ridley's death, and this is part of what makes Gunnar responsible for Ridley's death in the actual scenario. Again, the thought is that shooting someone can only make you responsible for that person's death if failing to shoot him would not have causally contributed to his death. For, if failing to shoot him would have causally contributed to his death, then it seems that you cannot be responsible for his death *in virtue of* having shot him.

These reflections motivate the following principle:

Failing to Cause Principle: *S* is responsible for *X* in virtue of having done *Y* only if, if *S* had not done *Y*, then *S* would have failed to bring about *X* by failing to do *Y*. Moreover, the fact that *S* would not have brought about *X* by failing to do *Y* helps ground *S*'s responsibility for *X* in the actual scenario.³⁸

For example, if Gunnar is responsible for Ridley's death in virtue of having shot him, then, if Gunnar had failed to shoot Ridley, his failure to shoot him would not have brought about Ridley's death. Moreover, the fact that his failure to shoot him would not have brought about Ridley's death helps ground Gunnar's responsibility for Ridley's death in the actual scenario. Similarly, if Jones is responsible for choosing to shoot Smith in virtue of not having considered the moral reasons not to shoot him, then, if Jones had considered those moral reasons, his considering those moral reasons would not have brought about his choice to shoot Smith. Moreover, the fact that his considering those moral reasons would not have brought about his choice to shoot Smith helps ground Jones's responsibility for his choice in the actual scenario.

The Failing to Cause Principle is a plausible principle of responsibility that states that certain possibilities where an agent fails to cause a relevant event (one that he actually causes) play a role in grounding the agent's responsibility. Hence it captures the idea that responsibility is importantly linked with the possibility to "fail to be involved." Since all the principle says is that an agent who is responsible for *X* in virtue of *Y* would not have caused *X* by not doing *Y*,

³⁸ In "Actuality and Responsibility," I argue for the truth of a principle along these lines, and I explain how it fits within a view of responsibility that is opposed to the alternative-possibilities model. I am understanding the principle in such a way that *Y* can be either an act or an omission: if *Y* is an act, then failing to do *Y* is an omission, and, if *Y* is an omission, then failing to do *Y* is an act.

without specifying that the agent must have had the ability not to do *Y*, the principle captures the idea that responsibility is linked in some way with the possibility to fail to be involved, without claiming that responsibility requires a genuine ability to fail to be involved.

Finally, note that the Failing to Cause Principle rests on the same metaphysical assumptions about causation on which AP* rests. In particular, it rests on the intransitivity of causation. The principle claims that, for example, if shooting someone makes you responsible for his death, it is because failing to shoot him would not have causally resulted in the death, even if it would have causally resulted in something that causally resulted in the death (such as the shooting by the backup assassin in the GUNNAR AND RIDLEY case). Again, it is an interesting fact that a plausible principle of responsibility like the Failing to Cause Principle relies on a metaphysical assumption about the nature of causation. For it suggests that we can gain insight on the concept of responsibility, and on the problem of freedom and determinism, by paying closer attention to the metaphysics of causation.

CAROLINA SARTORIO

University of Arizona