

No cover  
image  
available

## The Oxford Handbook of Moral Responsibility

Dana Kay Nelkin (ed.), Derk Pereboom (ed.)

<https://doi.org/10.1093/oxfordhb/9780190679309.001.0001>

Published: 2022

Online ISBN: 9780190679330

Print ISBN: 9780190679309

CHAPTER

# 17 Responsibility and Causation

Carolina Sartorio

<https://doi.org/10.1093/oxfordhb/9780190679309.013.8> Pages 348–C17.N\*

Published: 14 February 2022

## Abstract

This chapter discusses the various connections that exist between moral responsibility and causation. It focuses on the role played by causation in different forms of responsibility, including basic responsibility and derivative (or inherited) responsibility, responsibility for actions and for omissions, and responsibility as blameworthiness and as praiseworthiness. It also examines the phenomenon of causal overdetermination and the role played in those cases by the individual and/or collective contributions of morally responsible agents. Finally, it discusses the role played by causation in certain views of responsibility that are inspired by Harry Frankfurt's attack on the classical model of free will and responsibility based on alternative possibilities (and, in particular, in the corresponding "Frankfurt-style" scenarios). These are views according to which the free will of agents is exclusively a function of the actual causes of their behavior, or of the factors that actually explain how they acted.

**Keywords:** [moral responsibility](#), [causation](#), [basic responsibility](#), [derivative responsibility](#), [responsibility for omissions](#), [causation by omission](#), [causal overdetermination](#), [free will](#), [Frankfurt-style cases](#)

**Subject:** [Moral Philosophy](#), [Philosophy](#)

**Series:** [Oxford Handbooks](#)

## 1. Introduction

MORAL responsibility and causation are connected in important ways. Most obviously, we have the power to causally affect the world, and, arguably, it is thanks to those causal powers that we can be morally responsible (praiseworthy or blameworthy) for outcomes that we help bring about. Causation is, on this picture, our "bridge to the world." Thus causation seems central to the project of understanding the conditions under which we can be morally responsible for outcomes. This is a *derivative* form of responsibility. That is to say, our responsibility in these cases is not basic but derived or inherited from other things for which we are responsible, such as the choices that we made and that led to those outcomes.

Also, and a bit less obviously, causation is relevant to our moral responsibility in cases of *basic* responsibility, that is to say, in cases where our responsibility is not inherited from our responsibility for other things. Consider certain choices for which we may be basically responsible. At the very least, causation

matters in these cases in that the causal history of those choices (how we made those choices, or what led to our making them) is relevant to whether we made them freely. And freedom is a requirement for moral responsibility—at least for the form of responsibility that I will be concerned with here.<sup>1</sup>

p. 349 In this chapter I examine different aspects of the connection between causation and moral responsibility. I discuss derivative responsibility in section 2 and basic responsibility in section 3.

## 2. Derivative Responsibility and Causation

It is natural to think of causation as the vehicle of transmission of responsibility (Sartorio 2007). That is to say, assuming that we have figured out what kinds of things we are basically or non-derivatively responsible for, it is natural to think of causation as the relation that allows for responsibility to be transmitted or passed on to other things, things for which we can then be derivatively responsible (assuming we also meet the relevant epistemic conditions for responsibility; more on this later).

In this section I discuss this picture in more detail. As we will see, the connection between causation and responsibility turns out to be quite complex, even when we are dealing with derivative responsibility (where, as I noted before, the link between the two concepts is most obvious). In what follows I discuss some of the main issues that arise when thinking about derivative responsibility and the connection with causation.

The role that causation plays in transmitting moral responsibility can be captured by means of principles of derivative responsibility. In general terms, these are principles that give conditions (necessary and/or sufficient conditions) for the transmission of responsibility from more basic things to outcomes, where one of the conditions is a causal condition. Here are a couple of examples of principles of that kind that seem at least initially plausible:

**N** (for “Necessary”): Responsibility for an outcome requires responsibility for behaving in a way that caused the outcome.

**S** (for “Sufficient”): If an agent is responsible for an act, the act caused an outcome, and the agent foresaw that the act would cause (or was likely to cause) that outcome, then the agent is also responsible for the outcome.

Note that these principles make different kinds of claims about derivative responsibility: N identifies a potentially necessary condition for that kind of responsibility and S identifies a potentially sufficient condition for that kind of responsibility. But both principles involve a causal condition: In the case of N, the causal condition itself is claimed to be a necessary condition for responsibility, and, in the case of S, the causal condition is conceived as one in a larger set of conditions that are only jointly sufficient for responsibility.

p. 350 Note that N and S could be combined into a stronger principle stating conditions that are *both* necessary and sufficient for derivative responsibility. Here is a possible formulation of such a principle:

**N&S**: An agent is responsible for an outcome if and only if there is some behavior of hers such that: she is responsible for that behavior, the behavior caused the outcome, and she foresaw that the behavior would cause (or was likely to cause) that outcome.

Given that the stronger principle N&S entails both N and S, it inherits any problems that N or S might have. So here I will limit the discussion to an analysis of potential problems that arise for each of the individual principles, N and S.

I'll start with S. Among other things, S attempts to capture the well-known fact that there are epistemic conditions that an agent needs to meet in order to be morally responsible for anything. In the case of responsibility for outcomes, it is common to understand the epistemic condition in terms of the concept of foreseeability. The lack of foreseeability typically excuses. Imagine that an evil mastermind has secretly rigged things in such a way that eating the last apple in your refrigerator would trigger a series of events that would result in a remote village being wiped out. If you eat the apple and the outcome ensues, you fail to be responsible for the outcome because you didn't foresee that it would happen. Thus in this case S wouldn't entail that you are responsible for the outcome. In contrast, S would entail that you are responsible for the outcome in a case where the outcome was a foreseen result of what you did. Imagine, for example, that you drink heavily at a party knowing that you will have to drive back home intoxicated, and foreseeing the likely consequences of your decision. Imagine that on your way back you are involved in an accident where somebody is hurt as a result of your driving under the influence of alcohol. In this case S correctly entails that you are responsible for the harm that you caused. So S seems to capture the intuitive idea that behaving in a way that causes harm is enough for you to be derivatively responsible for the harm, in cases where you are responsible for that behavior and you foresaw that the harm was likely to result from it.

Now, note that, as it stands, S won't be able to capture other cases where we want to hold agents morally responsible for outcomes, such as some cases of negligence. Imagine that you act carelessly, in a way that a reasonable person wouldn't have, but without actually foreseeing that you could cause any harm. In some such cases we probably would want to hold you responsible for the outcome, even if you didn't foresee the harm, for (we want to say) you *should* have foreseen it. So, if we want S to capture these cases, this means that the epistemic condition would have to be relaxed a bit. Recall that S only gives a sufficient condition for responsibility, so negligence cases are not strictly speaking counterexamples to S (they are just cases where S is silent on the agent's responsibility). But if we are looking for a single unifying principle that covers both kinds of cases, cases of foreseen harms and cases of negligent harms, then the current formulation of S won't do.

p. 351 Note that accepting that agents can be responsible in some cases of unforeseen harms lowers the bar considerably for the epistemic conditions that agents would have to meet in order to be responsible. As a result, some argue that it is only appropriate to hold agents responsible in cases of this kind if the agents' lack of foresight can be traced back to some earlier decision for which they are responsible (see, e.g., Zimmermann 1986). This is sometimes referred to as a "tracing" condition. In turn, others disagree, arguing that agents can be responsible in some of these cases even where their lack of foresight cannot be traced to an earlier event for which they are responsible (see, e.g., Sher 2009 and Clarke 2014). (For a more comprehensive discussion of these and related issues, see the collection of papers in Nelkin and Rickless 2017.)

Another potential problem with S concerns cases of causal deviance. These are cases where the agent's behavior results in the harm, but in a way that radically departs from the way in which it was expected or could have reasonably been expected to do so. Imagine that an evil husband gives his wife a glass of water to drink, mistakenly thinking that it contains poison. Imagine that when the wife drinks the water she unexpectedly chokes on the water and dies. Here it seems that we can't hold the evil husband responsible for his wife's death, even if the death was the foreseen result of an act for which he is responsible. The reason is that the causal chain is too "deviant": there is a significant departure from the way in which the act was expected to issue in the death (for a discussion of the deviance condition, see, e.g., Feinberg 1970).

In order to deal with this problem, one could add an extra condition requiring that the causal chain linking the behavior to the outcome not be deviant, or one could tweak one of the existing conditions in such a way that deviant causal chains are ruled out. For example, one could revise the epistemic condition to say that the agent foresaw (or should have foreseen) that the harm would come about *in roughly the way it did*. As

usual, however, it is hard to give a precise account of what a deviant causal chain is, or of how we should understand “roughly” in this revised formulation of the epistemic condition.

There are other issues with S that concern the two appearances of the word “responsible.” Recall S’s formulation:

**S:** If an agent is responsible for an act, the act caused an outcome, and the agent foresaw that the act would cause (or was likely to cause) that outcome, then the agent is also responsible for the outcome.

Imagine that the outcome is bad. Then the question arises: Should we just understand “responsible” as “blameworthy”? If so, is this for one or both occurrences of the word?

In order to answer these questions, consider the following scenario. You need to leave a room or you’ll asphyxiate. You know that opening the door would result in an explosion in an adjacent room and cause some damage. You freely choose to open the door and walk out of the room. As predicted, the explosion occurs and the harm is done. In this case we don’t want to hold you blameworthy for the harm, because you had a good excuse: Your life was at stake. However, note that there is clearly a sense (a more morally neutral sense) in which you are responsible for your act of opening the door. After all, you did it freely, you were aware of what you were doing, and of the consequences it would have. Plus, your act in fact resulted in the expected outcome (in the expected way). Thus, if we understand the first occurrence of the word “responsible” in principle S in this more morally neutral sense, and the second occurrence as “blameworthy,” S wrongly entails that you are blameworthy for the outcome in this case.

One way to address this issue is to understand both occurrences of the word “responsible” in the same way, for example, as meaning “blameworthy.” Here is a possible formulation of the revised principle:

**S (Blameworthiness):** If an agent is *blameworthy* for an act, the act causally resulted in a bad outcome, and in roughly the expected way, then the agent is also *blameworthy* for the outcome.<sup>2</sup>

Note that this principle doesn’t entail that you are blameworthy for the harm in the case where you open the door. For, given the special circumstances you were in, you are clearly not blameworthy for the act of opening the door in that case. Again, you are arguably responsible for that act (and for the outcome) in a more morally neutral sense of the word “responsible,” but you are not responsible for that act (or for the outcome) in the sense of being blameworthy.<sup>3</sup>

Note that we can formulate a similar principle about praiseworthiness:

**S (Praiseworthiness):** If an agent is *praiseworthy* for an act, the act causally resulted in a good outcome, and in roughly the expected way, then the agent is also *praiseworthy* for the outcome.

Arguably, this is also a better formulation than a principle that mixed praiseworthiness with a more neutral sense of responsibility.

To illustrate why, consider the following case discussed by Knobe (2003). The CEO of a company implements a policy only moved by the desire to increase the company’s profits. As it happens, a foreseeable side effect of the policy is that it will help the environment, but this is not what moves the CEO to act. In that case, the CEO doesn’t seem praiseworthy for helping the environment, despite the fact that he freely and knowingly chose to implement a policy that foreseeably resulted in such a benefit to the environment. But note that we also wouldn’t say that the CEO is *praiseworthy* for implementing the policy, precisely because he didn’t do it for the right reasons. (If he is responsible for implementing the policy, it is, again, only in a more morally neutral sense of responsibility, one that doesn’t involve praiseworthiness.) Thus, S

(Praiseworthiness) doesn't have the implication that the CEO is praiseworthy for helping the environment, which is the result we wanted.<sup>4</sup>

So far we have been looking at principle S and different reasons to think that it might need revision. Now let's switch to N, the principle that gives necessary conditions for derivative responsibility. Here is again principle N:

N: Responsibility for an outcome requires responsibility for behaving in a way that caused the outcome.

Like S, N is an intuitive principle: Roughly, it captures the idea that we can't be morally responsible for an outcome unless we are "part of the reason" it happened, or unless we were somehow involved in its causal history. Imagine that you tried to cause some harm and the harm happened, but it didn't happen because of what you did. In that case it seems that you cannot be morally responsible for the harm. Wanting or trying to be involved in the causing of harm isn't enough to make you responsible for the harm; in addition, you actually must have been involved in the causing of the harm. This is the thought that N is supposed to capture.

Despite its initial plausibility, N also faces some potential challenges. First, in principle it seems that we could be morally responsible for some non-causal consequences of our acts. Perhaps causation is our *first* bridge to the world, but then other non-causal notions of consequence can take over in such a way that we also end up being responsible for some non-causal results of our acts. As an example of how this could happen, note that, according to some views of causation, outcomes that are overly "extrinsic" in nature cannot enter in causal relations (see, e.g., Lewis 1986a). So imagine that an assassin kills the parents of a child, thus turning the child into an orphan. The child's becoming an orphan is a consequence of the assassin's act, but, on those views of causation, it is not a causal consequence of the assassin's act. For it is an overly extrinsic event: Its occurrence heavily depends on the obtaining of external circumstances (whether the child is an orphan is determined by what happens, not to him, but to his ↘ parents). On these views, then, when the assassin shoots the parents, this has some causal consequences (the parents' death) as well as some non-causal consequences that follow from those (the child's being an orphan). And, of course, in this case we want to blame the assassin for both consequences of his action. This suggests that we might have to revise N. Ultimately, whether it needs revision is something that will likely depend on what the true theory of causation (or the "causal relata") is, in particular, it will depend on whether overly extrinsic events can be causal consequences of acts.

p. 354

Also, consider the fact that we can be morally responsible for the outcomes of our *omissions*, in addition to the outcomes of our positive acts. For example, a parent who fails to attend to her child's needs can be responsible for the ensuing harm to her child. But some believe that omissions cannot be causes, for they think of omissions as absences (absences of actions of a certain kind), and they think that only (positive) events can be causes and effects (see, e.g., Dowe 2001 and Beebe 2004). If this is right, it follows that causation needn't even be our first bridge to the world: In the case of omissions, the consequences of our omissions are never causal, and N would have to be revised accordingly to accommodate this fact.

Others believe that omissions and absences in general can be causes (see, e.g., Schaffer 2000, Lewis 2004, and McGrath 2005). So, again, N's truth will hinge on the relevant facts about causation and the causal relata. But it is important to note that, even those who believe that omissions cannot be causes typically tend to assume that there is an alternative relation or concept that can ground our moral responsibility in those cases. For example, Dowe (2001) introduces the concept of "quasi-causation" that applies to omissions. Quasi-causation is, essentially, merely possible causation, in that it makes reference to causal relations that obtain in other possible worlds. For instance, a mother's neglecting her child may be a quasi-cause of the harm to the child in the sense that, had she attended to her child's needs, her doing so *would*

have caused the child's well-being. With this notion of quasi-causation in place, then, one could revise N to accommodate the outcomes of our omissions by requiring that the agent's behavior be a cause or a quasi-cause of the outcome. It is interesting to note, however, that, at least in the case of theories like Dowe's, causation still plays an important role in grounding the moral responsibility of agents by omission. Even though actual causation isn't what plays that role, *possible* causation is. (Another related proposal is Beebe's suggestion that omissions can causally explain without being causes of what they explain. Her account also appeals to possible causal relations; see Beebe 2004.)

p. 355 Finally, even setting omissions aside, N may be open to counterexample. For instance, consider cases of (symmetric) overdetermination, such as firing squad cases where several shooters kill a victim and each of them would have been sufficient to bring about the death. According to some views, overdeterminers are not causes (for discussion, see, e.g., Lewis 1986b, Postscript E). But presumably we want to blame the shooters (all of them) for the outcome. Conversely, there are views according to which overdeterminers *are* causes (see, e.g., Schaffer 2003); in fact, this is probably the majority view nowadays. So how we approach this will, once again, depend on our views about causation. But, here too, it is important to note that even those who assume that overdeterminers aren't causes typically think that there is another way to link the agents to the outcome in overdetermination cases. The obvious alternative is to appeal to some form of *collective* causation. In particular, Lewis (1986b, Postscript E) argues that the mereological sum of the overdetermining events is a cause of the outcome in overdetermination cases. So, in the firing squad case, the mereological sum of all the shootings causes the victim's death. If so, N could be revised by requiring that the agent be responsible for *some cause* of the outcome, which could be an individual behavior of hers or perhaps, in some cases, a more complex event involving her behavior and also the behavior of other agents. (Sartorio (2004 and 2015) argues that another form of collective causation, a disjunctive form, is required to deal with the responsibility of agents in other kinds of special scenarios. If so, this calls for a similar revision of N.)

Putting this all together, N may have to be replaced by something like the following (much more complex but probably more accurate and less vulnerable to counterexample) principle:

N\*: Responsibility for an outcome requires responsibility for some cause or quasi-cause of the outcome, or for something that had the outcome as a non-causal consequence.

In sum, we have seen that, although when it comes to derivative responsibility there is clearly an important connection between responsibility and causation, it is hard to say exactly what that connection is, or to express that connection in the form of completely general principles that take into account all the subtleties involved. In particular, as we have seen, in several instances the view that we should embrace on these issues will depend on what we take the relevant facts about the causal relation to be. Still, the conception of causation as, roughly, a *ground of transmission of responsibility* remains as a powerful idea, and one that seems basically right when it comes to responsibility for outcomes in the world.

### 3. Basic Responsibility and Causation

Now let us turn to the relation between causation and basic responsibility. As I anticipated in the introduction, one way in which causation matters to basic responsibility is that the causal histories of our acts arguably matter to whether we act freely, and thus to whether we are morally responsible for those acts. For example, at least part of what determines whether an agent made a choice freely is whether it was the causal result of coercion, compulsion, or an ordinary process of deliberation. But are actual causal histories *the only thing* that matters to freedom? This is less clear.



According to the classical model of freedom, this is not all that matters, for something else is required in addition to a certain kind of causal history, namely, having alternative possibilities of action. The thought is intuitive enough: we cannot act freely unless there is more than one way we could have acted. Thus, according to this view, in order to determine whether an act is free, we must look partly outside the actual causal history of events and into facts that pertain to whether the agent had the ability to do otherwise. On some views, this is understood in terms of purely counterfactual possibilities that were still within the agent's reach. But this classical view of freedom has been forcefully contested in recent years, since the publication of a highly influential paper by Frankfurt (1969). Although Frankfurt's main goal in that paper was to undermine the view that freedom requires alternative possibilities, some of the ideas put forth there also help motivate a simpler conception of freedom, one according to which freedom is *just* a function of actual causal histories. According to this view, all it takes for an act to be free is to be caused in a certain kind of way. As a result, it is natural to expect that causation will play a particularly central role in this conception of freedom (more on this later).

The view that freedom is exclusively a function of actual causal histories is supported by the same examples that Frankfurt used to undermine the idea that freedom requires alternative possibilities, which have since then been known as "Frankfurt-style" cases. Here is one of Frankfurt's own examples:

**Jones and the Neuroscientist:** A neuroscientist wants Jones to perform a certain action. He is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (he is an excellent judge of such things) that Jones is going to decide to do something other than what he wants him to do. If it were to become clear that Jones is going to decide to do something else, the neuroscientist would take effective steps to ensure that Jones decides to do what he wants him to do, by directly manipulating the relevant processes in Jones's brain. As it turns out, he never has to show his hand because Jones, for reasons of his own, decides to perform the very action he wants him to perform. (Frankfurt 1969)

As Frankfurt noted, Jones seems to act freely and be responsible for his act, although, given the presence of the neuroscientist, he lacked alternative possibilities. So examples of this kind can be used to argue against the idea that freedom requires alternative possibilities. (Although whether they are actually successful in undermining the classical conception of freedom is still a matter of controversy. The literature on Frankfurt-style cases is now huge; a good resource is the collection of papers in Widerker and McKenna 2003.)

At the same time, cases of this kind can be used to motivate a different picture of freedom, one according to which all that matters to freedom are the actual causal histories. Briefly, here is why (for a more extended discussion of this point, see Sartorio 2017). As Frankfurt noted, these examples illustrate an important distinction: The distinction between the factors that make it the case that the agent couldn't have done otherwise, on the one hand, and the factors that explain why the agent acted as he did, on the other. I'll use the labels "inevitability factors" to refer to factors of the first kind and "explanatory factors" to refer to factors of the second kind. Frankfurt thought that examples like Jones and the Neuroscientist help us see that the only factors that are relevant to freedom are the explanatory factors. First, note that in a case like Jones and the Neuroscientist the inevitability factors and the explanatory factors come apart. The explanatory factors are the factors involved in Jones's own process of deliberation that resulted in his choice, and the inevitability factors are the factors concerning the neuroscientist's existence that make it the case that Jones couldn't have made any other choice. Second, given that the neuroscientist never intervenes in the actual course of events, many of us tend to regard his presence as simply irrelevant to Jones's responsibility for his choice. Intuitively, Jones seems just as responsible as he would have been if the neuroscientist hadn't been present, given that he made his choice completely on his own, on the basis of his

own reasons. So, when the inevitability factors and explanatory factors come apart, we see that only the explanatory factors are relevant to the agent's freedom.

In contrast, in typical cases where an agent is unable to do otherwise and is intuitively not responsible for what he does, the inevitability factors coincide with the explanatory factors. For example, in a case where an agent is coerced to act in a certain way, the coercive threat is what makes the act inevitable, but it is *also* what explains why the agent did what he did. As a result, in these cases it is not immediately obvious what accounts for the agent's lack of responsibility, if it is the lack of alternatives or just the particular kind of causal history that the act has. For this reason, Frankfurt thinks, cases like Jones and the Neuroscientist put us in a better position to decide this issue (and, on his view, they support the idea that only the explanatory factors are relevant).

Of course, this is not the place to try to decide whether Frankfurt is right about this. So my goal in what follows will be a much more modest one. I will look more closely into this alternative picture of freedom (the one inspired by Frankfurt-style cases and motivated by Frankfurt's remarks on the distinction between inevitability factors and explanatory factors), and I will examine a number of important questions that arise for that view. Given the topic of this chapter, I am particularly interested in discussing the role that causation plays in a view of this kind.

The first question that arises for the view is this: How, exactly, should we understand the claim that *only the explanatory factors are relevant to freedom*? At first blush, the most natural way to understand it seems to be this:

**Actual Causes:** Only facts about the actual causes of an act (or some subset of those actual causes) are relevant to whether the act was done freely.

That is to say, Actual Causes claims that, when trying to determine whether an act was done freely, all we have to look at are those factors that were part of the causal history of the act. Recall that our focus now is basic responsibility. So imagine that an agent is basically responsible for the act of making a certain choice (a mental act) at a certain time. Then Actual Causes says that what accounts for the fact that the agent made that choice freely is only facts concerning the factors that actually caused the agent's choice. (Or, more precisely, some relevant subset of those factors. This qualification is warranted because acts, like all other events, have a great number of causes, and many of them ↪ probably won't be relevant to whether the agent acted freely. I'll leave this qualification implicit in what follows.)

p. 358

Actual Causes is a very simple and elegant view of freedom. Again, it is motivated by a natural way of understanding the thought that only the explanatory factors, the factors that actually explain an agent's behavior, are relevant to the agent's freedom. Causation plays a central role in a view of this kind. For note that what Actual Causes says is that freedom is exclusively grounded in certain *causal facts*. For example, what grounds the fact that an agent made a choice freely is the fact that the choice was caused by a process of a certain kind, say, an ordinary process of deliberation. Notice that this isn't just to say that there *was* such a process, but, more precisely, that such a process *causally resulted in* the agent's choice. Thus, on this view, causation plays an important role in that only causal facts (facts of the kind "C caused E") ground the freedom of agents.

However, Actual Causes may need some tweaking. First, imagine that one believes, again, that omissions cannot enter in causal relations (because one believes that omissions are absences and that absences in general cannot enter in causal relations, as discussed in the previous section). Imagine that one also believes that we can be basically responsible for some of our omissions, for example, for *not* making a certain choice at a certain time (the omission of a mental act). If omissions cannot enter in causal relations, then this means that omissions don't have causal histories, and thus Actual Causes entails that nothing is ever



relevant to whether we freely omit to act. So, depending, again, on one's view about the causal powers of omissions, one might need to revise Actual Causes in order to accommodate basic responsibility for omissions. And the natural solution to this problem would be, again, to broaden the view in whichever way is needed to accommodate omissions, by incorporating concepts that one's metaphysics of omissions allows for and that can plausibly ground the agent's moral responsibility in those cases (such as the concept of quasi-causation discussed in the previous section, or the idea that omissions can be causally explanatory without themselves being causes).

But there is a second reason to think that Actual Causes needs tweaking. It is the fact that there could be factors that help determine the constitution of the actual causal history in each case but that are not, themselves, part of the actual causal history (Sartorio 2011 and 2016). Unless one believes that causation is a primitive metaphysical relation, there will always be other more metaphysically basic facts that ground the relevant causal facts, and that are not themselves part of the actual causal history. Still, given that those facts help determine the constitution of the causal history, and given that the causal history is relevant to the agent's freedom, those facts seem to be indirectly relevant to the agent's freedom (they are relevant to the agent's freedom, but only given the role they play in determining the constitution of the causal history). However, as it stands, Actual Causes doesn't allow for the relevance of any such facts.

p. 359 For example, imagine that Ann is basically responsible for making the choice to lie to Bert at a certain time. Imagine that Ann's choice was caused by an ordinary process of deliberation, one that included weighing the reasons for and against lying to Bert, etc. On the view that we are considering, the fact that Ann's choice was caused in that way (let's call this fact the "causal history fact") accounts for the fact that it was made freely  $\hookrightarrow$  (the "freedom fact"). But imagine that one also believes in a reductive theory of causation, say, a *counterfactual* theory of causation, according to which causation reduces to counterfactual dependencies of some kind—dependencies of the sort "if event X hadn't occurred, then event Y also wouldn't have occurred."<sup>5</sup> On a view of this kind, the causal history fact in question is in turn grounded in a "counterfactual dependence fact" (a fact involving the relevant relations of counterfactual dependence), as represented by the following diagram:

Counterfactual dependence fact  $\rightarrow$  Causal history fact  $\rightarrow$  Freedom fact

As a result, on this picture, the freedom fact seems to be grounded in both the causal history fact and the counterfactual dependence fact. For it is implausible to suggest that only the causal history fact grounds the freedom fact, if the causal history fact is in turn grounded in the counterfactual dependence fact.<sup>6</sup>

So this means that Actual Causes might need to be revised to allow for the relevance of the grounding facts. But, once again, we see that *whether* it needs to be revised, and *how* we should revise it, ultimately depends on the relevant metaphysical facts about causation.

The preceding discussion suggests that the original principle, Actual Causes, might need to be replaced by a more complex version, perhaps one that looks like this:

**Actual Causes\*:** Only facts about the actual causes or quasi-causes of an act, and the grounds of those facts, are relevant to whether the act was done freely.

Now, at this point one might have a different kind of concern. One might worry that the resulting view of freedom will end up collapsing into a much more liberal model where *all* sorts of facts, not just facts about the actual causal history, play a role in grounding freedom. Perhaps it will even end up collapsing back into the classical model, according to which freedom is grounded in alternative possibilities! This would be an odd and unexpected result, since the two models are supposed to be rival conceptions of freedom.

p. 360

But this worry is unfounded. According to Actual Causes\*, the *only* facts that are relevant to freedom are still those facts that, directly or indirectly, concern the actual causal history. This means that all other facts are irrelevant to freedom, and this is likely to include facts about alternative possibilities. Still, it is interesting to note that it is in principle *possible* for a view to respect Actual Causes\* and to also be an alternative-possibilities view (see Sartorio 2016, ch. 1). What would have to happen for this to be the case is that the same facts that ground the causal history facts also make it the case that ↵ the agent has or lacks alternative possibilities. But notice that any such view would be an extremely watered-down version of the alternative-possibilities view. It would be a view according to which alternative possibilities are relevant to freedom, but only because they help determine what the actual causal history of events is. Most likely, a view of this kind will rely on some implausible assumptions about the metaphysics of causation. At any rate, this is not the kind of view that people typically have in mind when they think that freedom requires alternative possibilities of action: They don't typically have in mind facts that determine what the actual explanation of action is, but more robust facts that make it the case that the agent has genuine access to alternatives.

Finally, one might worry that Actual Causes\* is too simple in that facts pertaining to actual causal histories aren't rich enough, by themselves, to ground an agent's freedom. For example, it is common to suggest that an agent's freedom has to do with general abilities that can remain unexercised in the actual scenario, such as certain capacities for responsiveness to reasons. To the extent that those capacities are actually unexercised and are not relevant to the constitution of the actual causal history, Actual Causes\* would in fact entail that they are irrelevant to the agent's freedom.

In particular, it is interesting to note that most views that were inspired by Frankfurt's argument against the principle of alternative possibilities focus on some factors that don't pertain to the actual causal history. Notably, Fischer and Ravizza (1998) focus on certain modal properties of the actual causal mechanisms (the "responsiveness" of mechanisms), and a central thesis of their view is the claim that these properties are relevant to the agent's freedom despite their not being exercised in the actual scenario. (See also the view developed by McKenna (2013), which appeals to unexercised dispositions of agents.) Fischer and Ravizza seem to think that without appealing to something like the responsiveness of mechanisms we cannot fully capture everything that is involved in an agent's acting freely, in particular, we cannot capture the thought that a free agent is someone who is suitably responsive to reasons when he acts. Moreover, they argue that this is still within the framework of a conception of freedom that rejects the relevance of alternative possibilities, because a mechanism being suitably responsive doesn't require the existence of genuine alternative possibilities of action.

Now, although this may be true, the focus on factors that don't pertain to the actual causal history seems to be in tension with the initial motivation for the model of freedom based on actual causes. Frankfurt's insight, recall, was that only facts concerning the *actually explanatory* factors are relevant to an agent's freedom. Thus, in suggesting that other kinds of factors play a central role in grounding freedom, these views fail to respect that fundamental insight.<sup>7</sup>

p. 361

Alternatively, Sartorio (2016, ch. 4) argues that it is, in fact, possible to capture the idea that free agents are agents who are sufficiently responsive to reasons within a framework that respects the fundamental insight by Frankfurt. The key lies in realizing that causal histories are much richer than they seem to be at first sight. In particular, when agents act freely, the causal histories of their acts reflect their responsiveness to reasons because they contain, in addition to the reasons that motivate them to act, several *absences of reasons* that also explain their behavior. Intuitively, this is supposed to capture the fact that when we act we are not just responding to what *is* the case on the relevant occasion, but also to what is *not* the case on that occasion. All of these factors are part of the actual causal history or the actual explanation of the agent's behavior. On this view, even a very simple act such as my deciding to go for a walk one morning has a complex causal history, when it's done freely. The causal history includes the reasons that moved me to go

for a walk (say, the desire to exercise), but also the absence of reasons that would have moved me to stay home (say, the belief in an imminent snowstorm). (Although, here too, if absences cannot be causes, the claim about the role played by absences of reasons would have to be cashed out in different metaphysical terms. Once again, we see the relevance of the metaphysics of causation; for discussion of this point, see Sartorio 2016, ch. 2.)

## References

---

Beebe, H. (2004). "Causing and Nothingness." In J. Collins, N. Hall, and L.A. Paul (eds.), *Causation and Counterfactuals*. Cambridge, MA: MIT Press, 291–308.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Capes, J. (2012). "Blameworthiness Without Wrongdoing." *Pacific Philosophical Quarterly* 93(3): 417–437.

[Google Scholar](#) [WorldCat](#)

Clarke, R. (2014). *Omissions: Agency, Metaphysics, and Responsibility*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Dowe, P. (2001). "A Counterfactual Theory of Prevention and 'Causation' by Omission." *Australasian Journal of Philosophy* 79: 216–226.

[Google Scholar](#) [WorldCat](#)

Feinberg, J. (1970). "Collective Responsibility." In *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton, NJ: Princeton University Press, 222–251.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Fischer, J., and M. Ravizza (1998). *Responsibility and Control*. Cambridge: Cambridge University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Frankfurt, H. (1969). "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66(23): 829–839.

[Google Scholar](#) [WorldCat](#)

Frankfurt, H. (1971). "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68: 5–20.

[Google Scholar](#) [WorldCat](#)

Graham, P. (2011). "Fischer on Blameworthiness and 'Ought' Implies 'Can.'" *Social Theory and Practice* 37(1): 63–80.

[Google Scholar](#) [WorldCat](#)

Haji, I. (1998). *Moral Appraisability: Puzzles, Proposals and Perplexities*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Knobe, J. (2003). "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63: 190–193.

[Google Scholar](#) [WorldCat](#)

Lewis, D. (1986a). "Events." In *Philosophical Papers*, Volume II. New York: Oxford University Press, 241–269.

p. 362 Lewis, D. (1986b). "Causation." In *Philosophical Papers*, Volume II. New York: Oxford University Press, 159–213.

Lewis, D. (2004). "Void and Object." In J. Collins, N. Hall, and L.A. Paul (eds.), *Causation and Counterfactuals*. Cambridge, MA: MIT Press, 277–290.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

McGrath, S. (2005). "Causation by Omission: A Dilemma." *Philosophical Studies* 123: 125–148.

[Google Scholar](#) [WorldCat](#)

McKenna, M. (2013). "Reasons-Responsiveness, Agents and Mechanisms." In D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility*, Volume 1. Oxford: Oxford University Press, 151–183.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Nelkin, D., and S. Rickless (2017). *The Ethics and Law of Omissions*. New York: Oxford University Press.

[Google Scholar](#)   [Google Preview](#)   [WorldCat](#)   [COPAC](#)

Paul, L. A., and N. Hall (2013). *Causation: A User's Guide*. Oxford: Oxford University Press.

[Google Scholar](#)   [Google Preview](#)   [WorldCat](#)   [COPAC](#)

Raven, M. (2015). "Ground." *Philosophy Compass* 10(5): 322–333.

[Google Scholar](#)   [WorldCat](#)

Sartorio, C. (2004). "How to Be Responsible for Something Without Causing It." *Philosophical Perspectives* 18(1): 315–336.

[Google Scholar](#)   [WorldCat](#)

Sartorio, C. (2007). "Causation and Responsibility." *Philosophy Compass* 2(5): 749–765.

[Google Scholar](#)   [WorldCat](#)

Sartorio, C. (2011). Actuality and Responsibility. *Mind* 120 (480):1071–1097.

[Google Scholar](#)   [WorldCat](#)

Sartorio, C. (2015). "Resultant Luck and the Thirsty Traveler." *Method* 4(6): 153–172.

[Google Scholar](#)   [WorldCat](#)

Sartorio, C. (2016). *Causation and Free Will*. Oxford: Oxford University Press.

[Google Scholar](#)   [Google Preview](#)   [WorldCat](#)   [COPAC](#)

Sartorio, C. (2017). "Frankfurt-Style Examples." In M. Griffin, N. Levy, and K. Timpe (eds.), *Routledge Companion to Free Will*. London: Routledge, 179–190.

[Google Scholar](#)   [Google Preview](#)   [WorldCat](#)   [COPAC](#)

Schaffer, J. (2000). "Causation by Disconnection." *Philosophy of Science* 67: 285–300.

[Google Scholar](#)   [WorldCat](#)

Schaffer, J. (2003). "Overdetermining Causes." *Philosophical Studies* 114(1–2): 23–45.

[Google Scholar](#)   [WorldCat](#)

Sher, G. (2009). *Who Knew? Responsibility Without Awareness*. New York: Oxford University Press.

[Google Scholar](#)   [Google Preview](#)   [WorldCat](#)   [COPAC](#)

Widerker, D., and M. McKenna, eds. (2003). *Moral Responsibility and Alternate Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate.

[Google Scholar](#)   [Google Preview](#)   [WorldCat](#)   [COPAC](#)

Zimmerman, M. (1986). "Negligence and Moral Responsibility." *Noûs* 20: 199–218.

[Google Scholar](#)   [WorldCat](#)

Zimmerman, M. (1997). "A Plea for Accuses." *American Philosophical Quarterly* 34(2): 229–243.

[Google Scholar](#)   [WorldCat](#)

## Notes

---

- 1 This form of responsibility, which is sometimes known as "accountability," involves being eligible to be held to account for what you do, which requires acting freely or being in control of what you do (see, e.g., Fischer and Ravizza 1998).
- 2 Feinberg (1970) and Sartorio (2016: 77) propose different versions of this principle.
- 3 The principle might still need further refinement. This time imagine that there are two doors, A and B, both of which would issue in the explosion, but you wrongly believe that only opening door A would cause the harm. Imagine that you

open that door wanting to cause the harm, and the harm ensues. Some would want to say that in this case you are blameworthy for opening door A. (This assumes that you can be blameworthy for acts that aren't wrong, since opening door A in these circumstances arguably isn't wrong. For a defense of this view, see, e.g., Zimmerman 1997, Haji 1998, Graham 2011, and Capes 2012.) It seems to me, however, that you are clearly *not* blameworthy for the harm that you caused. But, assuming that you are blameworthy for opening door A, S (Blameworthiness) entails that you are also blameworthy for the harm. So this would call for further revision of the principle. What seems to be going on in this case is that, even assuming that the agent is blameworthy for her act, and even if the act leads to the harm in the expected way, there is still a mismatch between the agent's beliefs (such as the belief that she is acting wrongly) and what is actually the case. So, if a revision were needed, it would have to require the absence of that kind of mismatch.

- 4 Thanks to Tim Kearl for help with this paragraph. Knobe compares this case with another case where the CEO implements a policy motivated, again, just by the desire to make a profit, but foreseeing that the policy will have a harmful effect on the environment. In that case, the CEO seems blameworthy for the harm to the environment. So the contrast between these cases suggests that the conditions for blameworthiness and praiseworthiness differ in ways that track the agent's intentions or motivations for action.
- 5 It is worth noting that no plausible view of causation will identify causation with simple counterfactual dependence, because the resulting view is clearly open to counterexamples involving overdetermination, preemption, etc. But there are other more sophisticated and more plausible views out there that still reduce causation to facts about counterfactual dependence; for discussion, see Paul and Hall 2013.
- 6 Although some think that there are exceptions to the transitivity of grounding, in typical cases transitivity holds. For discussion of the grounding relation and its properties, see, e.g., Raven 2015.
- 7 Interestingly, Frankfurt's own positive view (Frankfurt 1971) may end up having the same problem, at least on some interpretations of the view. Roughly, the view is that being morally responsible requires us to have a will with a certain hierarchical structure, one where our higher-order desires align with our causally efficacious first-order desires. But nothing in this view seems to require that those higher-order desires themselves also be causally efficacious. If so, this is in tension with the idea that only causally efficacious factors are relevant to freedom. (For discussion of this point, see Sartorio 2016, ch. 1, sec. 3.2.2.).