

YOU JUST DIDN'T CARE ENOUGH

QUALITY OF WILL, CAUSATION, AND BLAMEWORTHINESS FOR ACTIONS, OMISSIONS, AND OUTCOMES

Mattias Gunnemyr and Caroline Torpe Touborg

WE MAY ASK at least two different questions about blameworthiness: What makes someone blameworthy *for* something, and *how* blameworthy is she?¹ In this paper, we shall focus on the first of these two questions. In the case below, for illustration, our immediate reaction is that Suzy is blameworthy for breaking the window:

Solo Suzy: Suzy is walking down the street. When she reaches the big house on the corner, she stops and considers. She has an intense dislike for the elderly couple who lives in the house, and she has just had an idea: she is going to upset them by breaking their window on the first floor. She carefully selects a stone and hurls it toward the window. She feels a jolt of satisfaction when she hears the sound of breaking glass. Then she walks on as if nothing has happened.

In virtue of what is Suzy blameworthy for breaking the window in this case? More generally: What distinguishes cases where an agent is blameworthy for something—an action, omission, or outcome—from cases where she is not?²

We aim to develop a compatibilist answer to this question. In doing so, we shall draw on two important approaches in the literature: the quality-of-will approach and the actual-sequence approach.

- 1 These two questions correspond closely to Zimmerman's distinction between the *scope* of blameworthiness, i.e., what you are blameworthy *for*, and the *degree* of blameworthiness, i.e., *how* blameworthy you are ("Taking Luck Seriously"). Zimmerman uses this distinction to reject moral luck by arguing that while luck matters for what you are blameworthy *for*, it does not matter for your *degree* of blameworthiness. In this paper, we stay neutral on the question of moral luck and focus simply on understanding blameworthiness *for*.
- 2 A parallel question may be asked about praiseworthiness for actions, omissions, and outcomes. In the following, we set this aside.

The quality-of-will approach is based on Strawson's suggestion that blame is tied to the reactive attitudes, particularly resentment, and that those attitudes in turn respond to an agent's quality of will: "The reactive attitudes I have so far discussed are essentially reactions to the quality of others' wills towards us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern."³ This idea has been developed by Arpaly, Björnsson, McKenna, Wallace, and others.⁴ Proponents of the quality-of-will approach tend to focus on the question of *how* blameworthy an agent is (the exception is Björnsson). However, as we will argue, quality of will also fits naturally when we are thinking about blameworthiness *for*. That is, quality of will is a natural starting point when analyzing the conditions under which others are warranted in reacting negatively to us (with resentment, indignation, or the like) in virtue of what we have done or brought about.

The actual-sequence approach takes its inspiration mainly from Frankfurt-style cases. On this approach, what matters in determining whether an agent is blameworthy for an action, omission, or outcome is the actual causal sequence leading up to that action, omission, or outcome.⁵

Björnsson's account elegantly combines these two approaches.⁶ The basic idea of his account is:

Basic Idea: You are blameworthy for X —where X may be an action, omission, or outcome—just in case there is a time t such that your poor quality of will at t stands in the right causal-explanatory relation to X .

Exactly how to understand "poor quality of will" is a matter of debate. Strawson characterizes poor quality of will in terms of manifesting ill will, indifference, or lack of concern; Björnsson characterizes it as caring less than is required; and McKenna characterizes it as showing insufficient regard.⁷ In our formal

3 Strawson, "Freedom and Resentment," 15.

4 Arpaly, *Unprincipled Virtue*; Björnsson, "Explaining Away Epistemic Skepticism about Culpability" and "Explaining (Away) the Epistemic Condition on Moral Responsibility"; McKenna, *Conversation and Responsibility*; and Wallace, *Responsibility and the Moral Sentiments*. See also, e.g., Shoemaker, *Responsibility from the Margins*; Smith, "Responsibility for Attitudes"; Talbert, "Moral Competence, Moral Blame, and Protest"; and Watson, "Responsibility and the Limits of Evil."

5 For developments of this idea, see Fischer and Ravizza, *Responsibility and Control*; and Sartorio, *Causation and Free Will*.

6 See Björnsson, "Being Implicated," "Explaining Away Epistemic Skepticism about Culpability," and "Explaining (Away) the Epistemic Condition on Moral Responsibility."

7 Strawson, "Freedom and Resentment"; Björnsson, "Being Implicated," "Explaining Away Epistemic Skepticism about Culpability," and "Explaining (Away) the Epistemic Condition on Moral Responsibility"; and McKenna, *Conversation and Responsibility*.

definitions, we shall refer simply to poor quality of will, leaving it open precisely how this should be understood. In our discussions, though, we often adopt Björnsson's proposal and understand quality of will in terms of care. The reader is free to substitute their own preferred understanding.

In this paper, we present a new way to develop the Basic Idea. First, we argue that it needs to be refined in a number of ways (section 1). Next, we present an account of the relevant causal-explanatory relation (section 2) and finalize the account of when you are blameworthy for actions, omissions, and outcomes, testing it on a number of cases (section 3). Finally, we show that this account also gives the right verdict in Frankfurt-style cases (section 4) and in collective-harm cases (section 5).

1. DEVELOPING THE BASIC IDEA

The Basic Idea already gives the intuitively right verdict in paradigm cases of blameworthiness *for*, such as Solo Suzy, where the intuitive verdict is that Suzy is blameworthy for throwing the rock toward the window (an action) and for breaking the window (an outcome). Here, Suzy has a poor quality of will—she dislikes the elderly couple who lives in the house and wants to break their window in order to upset them. Furthermore, Suzy's poor quality of will just before she throws her rock stands in the right causal-explanatory relation both to her throwing the rock and to the breaking of the window: Suzy's poor quality of will causes/explains her throwing the rock and the breaking of the window. Thus, Suzy is blameworthy both for throwing the rock and for the breaking of the window.

In Solo Suzy, Suzy intentionally breaks the window. In other cases, however, you may be blameworthy for something even though you did not do it or bring it about intentionally.⁸ When you have a poor quality of will, you may forget things you should remember, you may fail to notice things, or neglect to consider them. Suppose, for example, that you do not care as you should and therefore forget your best friend's birthday. In that case, we think that you are blameworthy for forgetting the birthday—even though, of course, you did not

8 Voluntarists about moral responsibility such as Fischer and Ravizza (*Responsibility and Control*) and Rosen ("The Alethic Conception of Moral Responsibility") would not agree. According to them, voluntary control is a precondition on being blameworthy, and you do not have voluntary control over, e.g., forgetting something. Still, it seems that you *are* blameworthy for something in such cases. Here, voluntarists typically argue that you are blameworthy for some earlier action or decision that you did have voluntary control over (given that you also satisfy some epistemic condition), such as failing to add a note in your calendar about your friend's birthday. This is the *tracing* strategy. There are, however, problems with the tracing strategy (see, e.g., Smith, "Attitudes, Tracing, and Control").

do this intentionally.⁹ The Basic Idea easily captures this: you are blameworthy for forgetting the birthday, because your poor quality of will—your not caring enough—stands in the right causal-explanatory relation to your forgetting. For another example, suppose that you plan a weekend at the golf course with your colleagues, without even considering visiting your injured daughter at the hospital.¹⁰ Here too, the Basic Idea captures why you are blameworthy: you failed to even consider visiting your daughter because you did not care enough about her.

However, the Basic Idea needs a number of refinements. In the remainder of this section, we introduce these refinements gradually, motivated by a series of cases.¹¹

The first refinement is motivated by the following observation: when we blame someone for something, this seems to imply that what they are blamed for is *bad*. On its own, however, the Basic Idea delivers the result that you may be blameworthy for a *good* outcome, if it is caused/explained by your poor quality of will. A simple way to fix this is to add a further necessary condition to the Basic Idea: you are blameworthy for *X*—an action, omission, or outcome—only if *X* is bad. However, this is not quite right. First, it is at best difficult, and at worst impossible, to define what it is for something—an action, omission, or outcome—to be bad *tout court*. It seems much easier to make comparative judgments that an action or outcome is worse than some alternative. Second, there are cases where it seems that you can be blameworthy for making a negative difference, even though the outcome that happens does not seem bad as such. Suppose, for example, that Sally and Bob are cooking chili together. Bob is careful about following the recipe. Sally, on the other hand, is more attentive to her phone than to her cooking and fails to put in some of the ingredients. The chili still ends up being tasty, though not quite as tasty as it would have been with all the ingredients. In this case, we think it makes sense to say that Sally is blameworthy for the chili turning out as it did, even though this outcome is

9 Smith, “Responsibility for Attitudes.”

10 McKenna, *Conversation and Responsibility*.

11 In “Explaining Away Epistemic Skepticism about Culpability” and “Explaining (Away) the Epistemic Condition on Moral Responsibility,” Björnsson suggests that a development of the Basic Idea should incorporate both an evaluative dimension (the object of blame must be bad) and the requirement that the agent’s bad quality of will should explain the object of blame in a normal way. This allows his account to handle the kinds of cases we discuss in the following: in cases where the outcome is only comparatively bad, Björnsson would suggest that comparative badness is a way to be bad, and in Tragedy, Björnsson would suggest that the agent’s bad quality of will does not explain the runaway consequences in a normal way. The refinements we suggest in this section draw on Björnsson’s insights and implement them in a new way.

not bad: we are warranted in reacting negatively to her since the chili turned out as it did, rather than turning out even better.

Both considerations point toward the same solution: that blameworthiness *for* involves a comparative element. Fully spelled out, you are not simply blameworthy for X , where X is some action, omission, or outcome. Rather, you are blameworthy for the occurrence of X rather than X^* , where X is worse than X^* .¹² This yields the following refined version of the Basic Idea:

Blameworthiness For #1: You are blameworthy for X rather than X^* just in case

1. X is worse than X^* , and
2. there is a time t such that your poor quality of will at t stands in the right causal-explanatory relation to X rather than X^* .¹³

This refined version easily handles the cases we have considered so far. However, problems still remain. Consider the following tragic variation of Solo Suzy:

Tragedy: Everything is as in Solo Suzy up to the point where the window breaks. But the consequences of the window breaking are dire. The husband is so upset at seeing the broken window that he suffers a heart attack and dies. Unable to cope with her husband's sudden death, the wife has a nervous breakdown and never fully recovers. Her daughter has to abandon a promising artistic career in Australia and come home to take care of her mother for the next several years. If Suzy had not broken the window, none of this would have happened. Instead, the couple would have continued to live happily together for many years, and their daughter would have been free to pursue her promising artistic career in Australia.

We have no doubt that Suzy is blameworthy for throwing her rock and for breaking the window. But is she also blameworthy for the runaway consequences: the husband's heart attack? the wife's nervous breakdown? the end

- 12 In ordinary language, we do not typically say that one is blameworthy for one thing *rather than another*. Instead, the relevant contrast is supplied by the context.
- 13 On the intended reading, X is an event that actually occurred, while X^* is a merely possible event that is incompatible with X . This is, for example, the case with the chili: the chili turning out as it did is an actual event, while the chili turning out even better is a merely possible event. Sally is blameworthy for the fact that the chili turned out as it did, rather than turning out even better. There is an alternative reading where both X and X^* are events that actually occurred. Suppose, for example, that Ben is also involved in the cooking and botches the dessert. If someone were to blame Sally for the failed dessert, we might correct them by saying, "Sally is blameworthy for the chili rather than (being blameworthy for) the dessert." The *rather-than* construction is ambiguous between these two readings. Throughout the following, we intend the first.

of the daughter's promising artistic career? We do not think so. According to Blameworthiness For #1, however, she is: Suzy's poor quality of will causes/explains both the breaking of the window and the unfortunate events that follow.

The case shows that there has to be a tighter fit between *what* an agent is blameworthy for and the *way* in which her quality of will is poor. What is the required fit? Here is a suggestion: an agent is blameworthy for X rather than X^* only if her poor quality of will *specifically in relation to X versus X^** stands in the right causal-explanatory relation to X rather than X^* . We may state the modified condition as follows:

Blameworthiness For #2: You are blameworthy for X rather than X^* just in case

1. X is worse than X^* , and
2. there is a time t such that your poor quality of will at t in relation to X versus X^* stands in the right causal-explanatory relation to X rather than X^* .

This captures what we need. To start with an easy case, suppose that although Suzy's quality of will is poor in that she wants to see the elderly couple upset, she still cares as she should about more serious outcomes, such as whether the elderly people might die or suffer a nervous breakdown, just as she still cares as she should about their daughter's artistic career. If she learned what happened next, she would be horrified and exclaim something like this: "It's true that I wanted to upset them, but I never wanted something like this to happen!" If we were to blame her, e.g., for the husband's death in this case, there clearly would not be the right fit between the *way* in which her quality of will was poor and *what* we blame her for: although her quality of will was poor in relation to the elderly couple's getting upset, it was *not* poor in relation to the possibility that the husband might die. Thus, condition 2 fails to be satisfied.

Blameworthiness For #2 successfully captures why Suzy is blameworthy for upsetting the elderly couple, but not for the husband's death, the wife's nervous breakdown, or the end of the daughter's promising artistic career. On further inspection, however, an unexpected difficulty arises: it is not actually clear that we still get the result that Suzy is blameworthy for throwing her rock or even for breaking the window. Consider Suzy's throwing her rock. According to Blameworthiness For #2, Suzy is blameworthy for throwing her rock rather than not only if there is a time when she has a poor quality of will in relation to throwing her rock rather than not. But as we have told the story so far, we have not said anything to the effect that Suzy has a poor quality of will in relation to throwing her rock rather than not—we have merely said that she

has a poor quality of will in relation to the elderly couple's getting upset. In that case, Blameworthiness For #2 does not entail that Suzy is blameworthy for throwing the rock.

Fortunately, there is an easy way to solve this difficulty. Even though Suzy's throwing her rock is not intrinsically worse than her not doing so, Suzy's throwing her rock *is* worse than not throwing in virtue of how her throwing (rather than not) is related to other things—such as the elderly couple's getting upset. Suzy's quality of will is poor in relation to her throwing in precisely this sense: she does not care as she should about some of the outcomes (such as the elderly couple's getting upset) that make her throwing worse than not throwing. We may capture this as follows:

Blameworthiness For #3: You are blameworthy for X rather than X^* just in case there is a Y and Y^* such that

1. X is worse than X^* , at least partly in virtue of Y being worse than Y^* , and
2. there is a time t such that your poor quality of will at t in relation to Y versus Y^* stands in the right causal-explanatory relation to X rather than X^* .

This secures the verdict that Suzy is blameworthy for throwing the rock: (1) Suzy's throwing the rock (X) is worse than her not throwing it (X^*), at least partly in virtue of the old couple's getting upset (Y) being worse than their not getting upset (Y^*); and (2) the time t just before she throws is such that Suzy's poor quality of will at t in relation to the elderly couple's getting upset (Y) versus not getting upset (Y^*) stands in the right causal-explanatory relation to her throwing the rock (X) rather than not (X^*). We similarly get the verdict that Suzy is blameworthy for breaking the window.

In cases where you do have a poor quality of will directly in relation to X versus X^* , we may set $Y = X$ and $Y^* = X^*$, effectively making Blameworthiness For #3 equivalent to Blameworthiness For #2, which is easier to use. In such cases, we will say that X *just is* worse than X^* (leaving it open in virtue of what X is worse than X^*). In this way, Blameworthiness For #3 straightforwardly gives the verdict that Suzy is blameworthy for the elderly couple's getting upset rather than not.

2. CHARACTERIZING THE RIGHT CAUSAL-EXPLANATORY RELATION

Until now, we have relied on an intuitive understanding of "the right causal-explanatory relation." In this section, we suggest that the relevant relation just is *causation*. The success of this kind of suggestion depends critically on the account of causation that is used. We consider this in detail and suggest that

the account of causation proposed by Touborg works well with our account of blameworthiness *for*.¹⁴

According to Touborg's account, there are two necessary and jointly sufficient conditions for causation. First, a cause has to *produce* its effect, in the sense that it has to be connected to its effect via a genuine process. Second, the effect has to *depend* on the cause, in the sense that the security of the effect has to depend on the cause.¹⁵

In the following, we first present the condition of production and then the condition of dependence. Fully spelled out, both conditions are complex; here we only include as much detail as we need to explain our account of blameworthiness *for*. For the sake of simplicity, we only consider causation in worlds with deterministic laws. Correspondingly, we assume determinism in the examples we consider below. However, we believe the account could be extended to also apply to causation in worlds with indeterministic laws.

2.1. *Production as Process-Connection*

Let us begin with the production condition. The guiding idea behind this condition is that a cause must be connected to its effect via a genuine process. This idea is familiar from the proposal that causation should be understood in terms of physical processes.¹⁶ In its simplest form, this proposal may be stated as follows:

Physical Process: *C* is a cause of *E* just in case *C* is connected to *E* via a physical process.

A physical process is here understood in terms of transfers of physical quantities—mass, energy, etc. To illustrate the idea, consider a paradigm case of causation, such as Suzy's throwing her rock and breaking the window. Here, there is indeed a physical process connecting Suzy's throw, through the trajectory of her rock and its impact on the window pane, to the shattering of the window.

14 See Touborg, *The Dual Nature of Causation*.

15 Touborg's account of causation is inspired by Hall's proposal that there are two concepts of causation: the concept of production and the concept of dependence. See Hall, "Two Concepts of Causation." Hall originally gave demanding conditions for production and dependence, and suggested that production and dependence were each individually sufficient for causation. However, he later abandoned this proposal in the face of counterexamples (Hall, "Structural Equations and Causation"). By contrast, Touborg suggests conditions for production and dependence that are much weaker, with production and dependence being individually necessary and jointly sufficient for a single concept of causation.

16 See, e.g., Dowe, *Physical Causation*.

However, trouble is not far to seek: the proposal that a cause must be connected to its effect via a physical process cannot accommodate omissions and absences as causes and effects. This means, for example, that it cannot deliver the intuitively correct verdict on a case like the following:

Indifferent John: John is walking along a beach and sees a child struggling in the water. John believes that he could save the child with very little effort, and in fact he could, but he is disinclined to expend any energy to help anyone. He decides not to save the child and continues to walk along the beach.¹⁷

Intuitively, John's failure to jump into the water and save the child is a cause of the child's death. However, John's failure to intervene does not transfer any physical quantities or exert any push or pull on the drowning child; it is a mere absence. Thus, Physical Process delivers the verdict that John's failure to jump into the water and save the child is *not* a cause of the child's death. This verdict is counterintuitive, and especially so in the context of blame.

The trouble extends further: even when the candidate cause and effect are both ordinary positive events, Physical Process delivers the verdict that there is no causal connection when an omission or absence features as an intermediary. Thus, proponents of Physical Process have to deny that pulling the trigger causes gunshot wounds, or that decapitation causes death, since there is an intermediary absence or omission in both cases: squeezing the trigger removes an obstacle that would have prevented the flight of the bullet; decapitation stops the blood flow, which would have prevented brain starvation.¹⁸ (Such cases are called "double prevention cases," since in these cases *C* causes *E* by preventing *D*, which would have prevented *E*.)

These cases show that it cannot be a necessary condition for causation that a cause must be connected to its effect via a physical process, when this is understood in terms of transfers of physical quantities. To capture the intuitive idea that some kind of connecting process is necessary for causation, we instead need a more abstract notion of a process, which can include omissions and absences. Touborg suggests that we may get such a more abstract notion of a process by starting from *minimal sufficiency*.¹⁹

Minimal sufficiency is a relation between a set of simultaneous events *S* and a later event *E*, where events are understood broadly, so as to include omissions

17 This case is based on Fischer and Ravizza's case "Sloth." See Fischer and Ravizza, *Responsibility and Control*, 125.

18 See, e.g., Schaffer, "Causation by Disconnection."

19 Touborg, *The Dual Nature of Causation*, ch. 6.

and absences. A set of simultaneous events S is minimally sufficient for a later event E just in case the occurrence of all the events in S guarantees (given the laws of nature) that E will occur; and if any event is removed from S , the remaining events no longer guarantee (given the laws of nature) that E will occur. Importantly, *minimal sufficiency* is a relation between actual events: only actual events—including actual omissions and absences—may feature in the set S ; and the later event E also has to be an actual event (where this includes actual omissions and absences).

Let us say that there is an *apparent process* from C to E when C is connected to E via a *chain* of such relations of minimal sufficiency. This is so when C belongs to a set of simultaneous events S_0 , which is minimally sufficient for some later event D_1 ; D_1 belongs to a set of simultaneous events S_1 , which is minimally sufficient for some later event D_2 ; ...; and D_n belongs to a set of simultaneous events S_n , which is minimally sufficient for the later event E . When we look more closely—by considering more and more intermediate times between C and E —we may sometimes find that the apparent process from C to E was not genuine: when we consider these intermediate times, we can no longer find a chain of relations of minimal sufficiency connecting C to E . In order for C to be *process-connected* to E , the connection must remain *when we consider more and more intermediate times* between C and E .²⁰

This notion of process-connection is sufficiently abstract to accommodate omissions and absences. Returning to the case of Indifferent John, for example, we find that John's poor quality of will (in relation to the child's drowning versus surviving) is process-connected to the child's drowning. John's poor quality of will at the time t just before he decides not to intervene belongs to a set of simultaneous events that is minimally sufficient for the child's drowning. And this connection remains no matter how many intermediate times we consider. Take, for example, the intermediate time t' after John has decided not to intervene and before the child has drowned. Here, we find that John's poor quality of will at t belongs to a set of simultaneous events that is minimally sufficient for his failure to intervene at t' (remember, his failure to intervene is an actual event), and his failure to intervene at t' in turn belongs to a set of simultaneous events that is minimally sufficient for the child's drowning. Thus, John's poor quality of will at t is process-connected to the child's drowning.

The notion of process-connection also allows us to distinguish genuine causes from preempted backups in cases such as the following:

20 The full definition of process-connection includes a further refinement: to be able to handle cases of late preemption, it makes use of a more demanding, *time-sensitive* relation of minimal sufficiency. For simplicity, we leave out this refinement. See Touborg, *The Dual Nature of Causation*, 143–48.

Backup Billy: Everything is as in Solo Suzy, except that Billy also wants the window to break. On seeing that Suzy throws her rock, Billy is satisfied and walks away. However, if Suzy had not thrown her rock, Billy would have thrown a rock himself a moment later, and the window would still have broken.

In this case, Suzy's poor quality of will guarantees that the window will break, and so does Billy's. However, only Suzy's poor quality of will (in relation to the elderly couple's getting upset versus not) is process-connected to the shattering of the window. To see this, the key is to look at intermediate times. Let t be the time just before Suzy throws her rock. Then Suzy's poor quality of will at t belongs to a set of simultaneous events that is minimally sufficient for the shattering of the window; and similarly, Billy's poor quality of will at t belongs to a set of simultaneous events that is minimally sufficient for the shattering of the window. However, when we bring in more and more intermediate times, we find that we can keep filling in the details in the chain connecting Suzy's poor quality of will to the shattering of the window—going from Suzy's poor quality of will, to her decision to throw, to her throwing the rock, to the rock's trajectory and impact on the window pane, and finally to the shattering of the window. By contrast, the connection between Billy's poor quality of will at t and the window shattering breaks down when we consider intermediate times. Consider, for example, a time t' after Suzy has thrown her rock and Billy has turned away, but before the window shatters. To connect Billy's poor quality of will to the breaking of the window, we would need an event D at this time t' —such as Billy's rock flying toward the window—such that Billy's poor quality of will belongs to a set of events that is minimally sufficient for D , and D in turn belongs to a set of events that is minimally sufficient for the window shattering. But there is no such event D in the actual world. For this reason, Billy's poor quality of will at t is not process-connected to the breaking of the window. This fits the judgment that Suzy's poor quality of will at t is a cause of the shattering of the window, while Billy's poor quality of will is not. In this way, the notion of process-connection does crucial work in distinguishing genuine causes from preempted backups.

However, process-connection is not sufficient for causation. The condition of process-connection needs to be supplemented with a second necessary condition for causation, requiring that a cause must make a difference to its effect. The need for this is brought out by the following three considerations.

First, process-connection on its own cannot yield the intuitively correct verdict on counterexamples to the transitivity of causation, such as the following switching case:

Trolley Trouble: Suzy is standing by a switch in the tracks as a trolley approaches in the distance. If she flips the switch, the trolley will travel down the left-hand track; if she does not flip the switch, it will travel down the right-hand track. Further ahead, the tracks converge again, and after that, five people are tied to the single track. Suzy wants the five to get run over, and she erroneously believes that they are tied to the left-hand track. She flips the switch so that the trolley travels down the left-hand track and subsequently runs over the five people. However, if she had not flipped the switch, the trolley would still have run over the five, reaching them via the right-hand track.²¹

Intuitively, Suzy's poor quality of will at t (the time just before she flips the switch) is not a cause of the five's death. However, Suzy's poor quality of will at t is process-connected to their death. Thus, process-connection is not sufficient for causation.

You might not immediately notice that Suzy's poor quality of will is process-connected to the death of the five. Indeed, Suzy's poor quality of will at t does not belong to a set of simultaneous events that is itself minimally sufficient for their death: a set that leaves out Suzy's poor quality of will and contains just the approach of the trolley, the layout of the tracks, etc., is sufficient for the trolley's running over the five. However, there is a *chain* connecting Suzy's poor quality of will at t to the death of the five: Suzy's poor quality of will at t belongs to a set of simultaneous events that is minimally sufficient for the trolley's journey along the left-hand track, and the trolley's journey along the left-hand track belongs to a set of simultaneous events that is minimally sufficient for the death of the five. This connection remains when we consider more intermediate times.

The problem arises since process-connection is a transitive relation—if C is process-connected to D , and D is process-connected to E , then C is process-connected to E . By contrast, causation is not transitive: it may happen that C is a cause of D , and D is a cause of E , but C is not a cause of E —as in *Trolley Trouble*.²² This is the first reason why the condition of process-connection needs to be supplemented.

21 This case is inspired by Foot, "The Problem of Abortion and the Doctrine of Double Effect"; Thomson, "Killing, Letting Die, and the Trolley Problem"; and Van Inwagen, "Ability and Responsibility." Switching cases like this are also common in the causation literature. See, e.g., Hall, "Structural Equations and Causation"; Paul and Hall, *Causation*, 232–37; and Sartorio, "Causes as Difference-Makers." Like we do here, Sartorio uses this kind of case to motivate the idea that a cause must make a difference to its effect, though her difference-making condition differs from ours.

22 For an overview, see Paul and Hall, *Causation*, ch. 5.

Second, process-connection cannot on its own accommodate *contrastive* causal claims. Process-connection is simply a relation between two actual events: an actual event C is process-connected to an actual event E . However, contrastive causal claims include merely possible events as contrasts to the cause C or the effect E , and the truth value of a contrastive claim depends on what these contrasts are.²³ The need to handle such contrastive causal claims is especially pressing when we are concerned with blameworthiness for actions, omissions, and outcomes: as we have seen above, Blameworthiness For is based precisely on a contrastive claim—namely that your poor quality of will stands in the right causal-explanatory relation to X rather than X^* .

Third, process-connection on its own cannot distinguish between causes and background conditions. Suppose, for example, that Selma has no royal connections. Is the queen of Sweden's failure to water Selma's flowers a cause of their death? Intuitively, it is not.²⁴ However, the queen's failure to water Selma's flowers *is* process-connected to their death. So if we take process-connection to be sufficient for causation, we cannot accommodate the intuitive verdict in this case.

These difficulties have a common solution: recognizing that there is a second necessary condition for causation, which captures the intuitive idea that a cause must make a difference to its effect.

2.2. Dependence and Security

The core idea that causes are difference-makers is familiar. For example, David Lewis writes that “we think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it.”²⁵ This idea is the starting point for counterfactual accounts of causation. In its contrastive form, it may be stated as follows:

Simple: Suppose that C occurs at t , E occurs later, and E^* is incompatible with E . Then C is a cause of E rather than E^* just in case, if C had not occurred, then E^* would have occurred instead of E .

The heart of Simple is the counterfactual: “if C had not occurred, then ...” To evaluate this counterfactual, we first identify all the worlds where C does not occur. Among these, we consider the worlds that are *closest* to the actual world

23 See, e.g., Schaffer, “Contrastive Causation” and “Causal Contextualism.”

24 See, e.g., Hart and Honoré, *Causation in the Law*, 38; and McGrath, “Causation by Omission.”

25 Lewis, “Causation,” 557.

@. If the consequent is true in each of these worlds, then the counterfactual is true; otherwise it is false.

The relevant notion of closeness is standardly understood in terms of similarity between entire worlds. Following Paul and Hall, we prefer instead to understand it in terms of similarity between states of worlds at times.²⁶ Thus, we shall say that two possible worlds w and w^* are close-at-time- t to the extent that the state of w at t is similar to the state of w^* at t . Supposing that C occurs at time t , this means that the counterfactual “if C had not occurred, then . . .” is true just in case the consequent is true in each of the closest-to-@-at- t worlds where C does not occur.

Even with this clarification, a question remains: What replaces C in the closest-to-@-at- t worlds where C does not occur? An obvious answer is that C is replaced by an event that is as similar as possible to C , without satisfying C 's conditions of occurrence. However, this proposal yields intuitively false results. As Lewis writes: “if C had not occurred and almost- C had occurred instead, very likely the effects of almost- C would have been much the same as the actual effects of C . So our causal counterfactual will not mean what we thought it meant, and it may well not have the truth-value we thought it had.”²⁷ That will not do. We need an alternative proposal about what replaces C .

Our preferred answer is that, when we evaluate counterfactuals, we do not in fact consider *all* possible worlds. Rather, we only consider a restricted class of possible worlds—namely, those possible worlds that we take to be relevant. This restricted class of possible worlds is itself a causal relatum; we shall call it a *possibility horizon*.²⁸ Our chosen possibility horizon will typically not contain any worlds where C is replaced by almost- C . Rather, it will typically contain only worlds where C either occurs or is replaced by a contextually salient alternative C^* that is quite different from C . When we only consider the worlds within such a possibility horizon, we find that in the *closest* worlds where C does not occur, it is replaced by C^* .

Based on this, we may now give the following more developed version of Simple:

*Simple**: Suppose that C occurs at time t , E occurs later, and E^* is incompatible with E . Then C is a cause of E rather than E^* within possibility horizon H just in case there is at least one world in H where C does not occur, and in the closest-to-@-at- t world(s) in H where C does not occur, E^* occurs instead of E .

26 Paul and Hall, *Causation*, 47–49.

27 Lewis, “Causation as Influence,” 90.

28 See Touborg, *The Dual Nature of Causation*, ch. 5.

Simple* can capture our intuitions in a wide range of cases. In particular, it successfully handles the cases that presented difficulties for process-connection. In Trolley Trouble, Simple* entails that Suzy's poor quality of will is not a cause of the five's death, because Suzy's poor quality of will makes no difference to their fate—they would have died either way. Furthermore, Simple* is tailor-made to handle contrastive causal claims, such as "Sally's not caring about the chili caused the chili to be just good, rather than excellent." Finally, Simple* captures the verdict that the queen of Sweden's failure to water Selma's flowers did not cause them to die: in ordinary contexts, it is not a relevant possibility that the queen waters Selma's flowers. Rather, the queen's failure to water the flowers is treated as a background condition. Thus, the possibility horizon that is in play in an ordinary context only contains worlds where the queen does not water the flowers and thus the requirement that "there is at least one world in H where C does not occur" fails to be satisfied.

As is well known, however, Simple* does not give a necessary condition for causation: there are cases where C is clearly a cause of E , even though E would still have occurred if C had not. We have already seen such a case: in Backup Billy, it is clear that Suzy's throwing her rock is a cause of the window shattering. However, if Suzy had not thrown her rock, the window would still have shattered—because, in that case, Billy would have thrown *his* rock. Cases such as this show that in order to capture the idea that making a difference is necessary for causation, we need a more subtle notion of difference-making: one that can capture, e.g., how Suzy's throw makes a difference to the shattering of the window, even though the window would still have shattered if she had not thrown.

The key to developing such a more subtle notion of difference-making is to pay attention to the *modal* features of events. In particular, when an event actually occurs, we may ask how easily it could have failed to occur; and when an event does not occur, we may ask how easily it *could* have occurred.²⁹ Touborg uses the notion of *security* to capture this.³⁰

On Touborg's account, whenever an event actually occurs, it has *positive security*. However, it may have a higher or lower *degree* of positive security. In some cases, an event E actually occurs, but when we consider what was the case at some earlier time t , we find that if things had been just slightly different at

29 More carefully: when a particular *type* of event does not occur, we may ask how easily an event of this *type* could have occurred. Speaking of types of events solves the difficulty that we cannot refer determinately to an event that did not occur. For simplicity, we suppress this complication in the text.

30 Here we use the simplified definition of security presented in Touborg, "Hasteners and Delayers." For more detail, see Touborg, *The Dual Nature of Causation*, ch. 8.

time t , E would not have occurred. In such cases, we shall say that E had a low degree of positive security at time t . Suppose, for example, that Suzy in Solo Suzy throws her rock toward the window and breaks it, but if there had been just a slight gust of wind at t (the time when Suzy threw her rock), a swaying branch would have deflected her rock, and the window would have remained intact. In this case, the breaking of the window has a very low degree of positive security at t . In other cases, an event E actually occurs, and when we consider what was the case at some earlier time t , we find that things would have had to be quite different at t in order for E not to occur. In such cases we shall say that E had a high degree of positive security at t .

Whenever an event fails to occur, this event has *negative security*. Once again, it may then have a higher or lower *degree* of negative security. Consider some event E that does not actually occur, and consider some time t prior to the time when E would have occurred, if it did occur. We may now ask: How different would things have to be at t in order for E to occur? If things would only have to be ever so slightly different at t in order for E to occur, we shall say that E has a low degree of negative security at t : although E does not happen, circumstances at t are such that it is *close* to happening. If, on the other hand, things would have to be quite different at t in order for E to happen, we shall say that E has a high degree of negative security at t : considering the circumstances at t , E is *far* from happening.

More formally, we may understand security-at-a-time in terms of the distance-at-a-time between worlds. We have already introduced the notion of distance-at-a-time above, when we discussed the evaluation of counterfactuals. As a reminder: two possible worlds w and w^* are close-at-time- t to the extent that the state of w at t is similar to the state of w^* at t . Based on this, we may define security-at-a-time as follows:

If an event E occurs in w , then E has positive security in w , and its degree of positive security at an earlier time t is given by the distance-at- t between w and the closest-to- w -at- t world(s) where E does not occur.

If an event E does not occur in w , then E has negative security in w , and its degree of negative security at an earlier time t is given by the distance-at- t between w and the closest-to- w -at- t world(s) where E occurs.

This notion of security allows us to capture a more subtle notion of difference-making: making a difference to the *security* of an event. A cause does not have to make a difference as to whether its effect occurs or not. But it *does* have to make a difference to the security of its effect: supposing that C occurs at time t , it has to be the case that if C had not occurred, E would have been *less secure*

at t than it actually was. In the case of contrastive causal claims, such as “ C is a cause of E rather than E^* ,” C has to make a difference to the security of both E and E^* : supposing again that C occurs at t , it has to be the case that if C had not occurred, E would have been *less secure* at t and E^* would have been *more secure* at t than what was actually the case.³¹ We shall call this kind of difference-making *security-dependence*.

2.3. Causation

So far, we have introduced two necessary conditions for causation: the condition of process-connection and the condition of security-dependence. Neither of these two conditions can stand alone. The condition of process-connection needs help from security-dependence when dealing with switching cases, contrastive causal claims, and the distinction between causes and background conditions; the condition of security-dependence needs help from process-connection when dealing with preemption cases such as Backup Billy. But together, these two conditions are jointly sufficient for causation, yielding the following account:³²

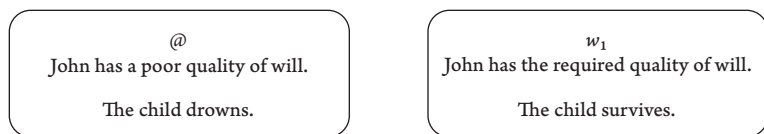
Causation: Suppose that C occurs at t , E occurs later, and E^* is incompatible with E . Then C is a cause of E rather than E^* within possibility horizon H just in case

- a. C is process-connected to E , and
- b. there is at least one world in H where C does not occur, and in the closest-to-@-at- t world(s) in H where C does not occur, E is *less secure* at t and E^* is *more secure* at t than they are in @.

This account of causation handles the cases we have considered so far.

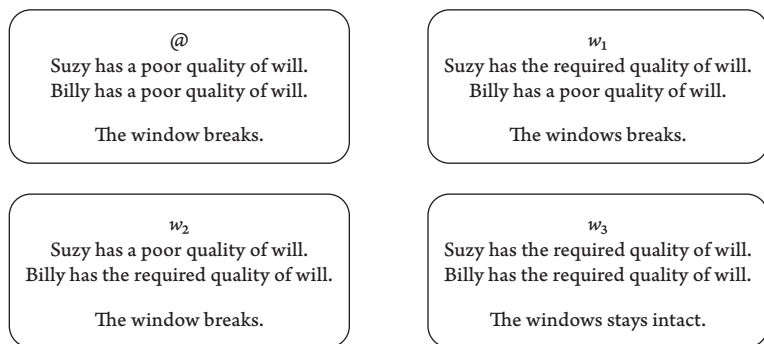
Consider first Indifferent John. As before, let t be the time just before John decides not to intervene. We have already seen that John's poor quality of will at t is process-connected to the child's drowning. We may now consider whether John's poor quality of will also satisfies the condition of security-dependence within the possibility horizon below:

- 31 The suggestion that a cause must make its effect more secure is somewhat similar to the controversial suggestion that a cause must raise the probability of its effect. In particular, (apparent) counterexamples to the suggestion that causes are probability-raisers can be translated into (apparent) counterexamples to the suggestion that causes make their effects more secure. However, the notion of security within a possibility horizon offers resources to resist such counterexamples. We therefore do not think they threaten the proposal. For discussion, see Gunnemyr, *Reasons, Blame, and Collective Harms*, 284–91; and Touborg, *The Dual Nature of Causation*, 239–43.
- 32 The account given here differs in one crucial respect from Touborg's account in *The Dual Nature of Causation*: Touborg does not include effect contrasts.

FIGURE 1 Possibility horizon H_j

Within H_j , the closest-to-@-at- t world where John does not have a poor quality of will at t is w_1 , where he has the required quality of will at t . Here, John jumps into the water and saves the child. Thus, the child's drowning has negative security at t in w_1 (since it does not occur in w_1) and positive security in @ at t (since it occurs in @). From this, it immediately follows that the child's drowning is *less secure* at t in w_1 than it is in @. Similarly, the child's survival is *more secure* at t in w_1 than it is in @. Thus, Causation yields the result that John's poor quality of will at t is a cause (within H_j) of the child's drowning rather than surviving. Therefore, John is blameworthy for the child's drowning rather than surviving.³³

Consider next Backup Billy. As before, let t be the time just before Suzy throws her rock. We may then consider what caused the window shattering within the following possibility horizon:

FIGURE 2 Possibility horizon H_b

The possibility horizon H_b includes a salient alternative to Suzy's poor quality of will (in relation to the elderly couple's getting upset versus not)—namely, her having the required quality of will; and it includes a salient alternative to Billy's poor quality of will—namely, his having the required quality of will. Independently

33 Is John also blameworthy for killing the child? Killing is sometimes understood simply as causing someone's death. If killing is understood in this way, then we would have to say that John killed the child. However, we think there are further conditions on killing (roughly related to the distinction between doing and allowing, and maybe with a proximate-cause requirement; see, e.g., Woollard, *Doing and Allowing Harm*), and John does not satisfy those further conditions—he merely allows the death of the child.

of this choice of possibility horizon, we have already seen that Suzy's poor quality of will at t is process-connected to the shattering of the window, while Billy's poor quality of will is not. We may now verify that Suzy's poor quality of will at t makes a difference to the security of the window shattering. The closest-to-@-at- t world within H_B where Suzy does not have a poor quality of will at t is w_1 , where she has the required quality of will at t . The window still shatters in w_1 , since Billy has a poor quality of will and therefore throws his rock when Suzy does not. However, the window shattering is *less secure* at t in w_1 than it is in @: compared with @, w_1 is *closer-at- t* to w_3 where the window does not shatter. In w_1 only one thing needs to change at t in order for the window not to break (namely, Billy's poor quality of will), whereas in the actual world @, two things need to change at t in order for the window not to break (namely, both Suzy's poor quality of will and Billy's poor quality of will). And similarly, the window's remaining intact is *more secure* at t in w_1 than it is in @. Thus, Causation yields the desired result: within possibility horizon H_B , Suzy's poor quality of will at t is a cause of the window's shattering rather than remaining intact, while Billy's poor quality of will is not.

Finally, consider Trolley Trouble. We have already seen that Suzy's poor quality of will at t (the time just before she flips the switch) is process-connected to the five's death. Now consider the possibility horizon H_T below, where the relevant alternative to Suzy's having a poor quality of will at t (in relation to the five) is her having the minimally required quality of will:



FIGURE 3 Possibility horizon H_T

Within this possibility horizon, Suzy's poor quality of will at t does not make any difference to the *security* of the five's death: there is no world where the five are not run over. Thus, their getting run over is infinitely secure, both in @ and in w_1 . And so, their getting run over is *just as secure* in w_1 as it is in @. We therefore find, as we should, that Suzy's poor quality of will is not a cause (within possibility horizon H_T) of the five's getting run over rather than not.

3. COMPLETING THE ACCOUNT OF BLAMEWORTHINESS FOR

We suggest that the causal-explanatory relation that has to hold between an agent's poor quality of will and what she is blameworthy for is causation, understood as suggested above.

As we have seen, causation is relativized to a possibility horizon. Thus, it may sometimes be the case that C is a cause of E rather than E^* within possibility horizon H_1 , while C is not a cause of E rather than E^* within a different possibility horizon H_2 . This feature of the general account of causation has a number of advantages. However, it would be unsatisfactory to say, e.g., that you are blameworthy for X rather than X^* within possibility horizon H_1 , but not within possibility horizon H_2 . We may avoid this relativity by insisting that what matters for blameworthiness is causation within the *relevant* possibility horizon. This raises a crucial question: What is the relevant possibility horizon when evaluating what an agent is blameworthy for?

Suppose we are evaluating whether your poor quality of will at time t (in relation to Y versus Y^*) is a cause of X rather than X^* , and that the purpose of this evaluation is to determine whether you are blameworthy for X rather than X^* . To make this evaluation, we start from the actual state of the world at time t . We then identify relevant alternatives to the way things were at time t . If you had a poor quality of will at t (in relation to Y versus Y^*), we think it is relevant that you could instead have had the quality of will (in relation to Y versus Y^*) that you were minimally required to have.³⁴ By contrast, it is not relevant that you could have had an even worse quality of will or a saintly quality of will far above what was minimally required. Similarly, if someone else had a poor quality of will at t , we think it is relevant that *they* could have had the quality of will they were minimally required to have. But again, it is not a relevant possibility that they could have had an even worse quality of will or a saintly quality of will. Other changes to what actually happened at t may or may not be relevant as well. This gives a criterion for determining which possible worlds belong to the relevant possibility horizon: if a possible world w represents a relevant alternative to how things were at time t and evolves forward in accordance with the laws of nature, then it is included in the relevant possibility horizon. Otherwise not. We may summarize this in the following rule of thumb.³⁵

Relevant Possibilities for Blame: To determine, for the purpose of attributing blame, whether your poor quality of will at time t (in relation to Y versus Y^*) is a cause of a later event X rather than X^* , it is a relevant

34 This proposal is closely related to Björnsson's proposal that what matters is how your quality of will falls short of what could be demanded. See Björnsson, "Explaining Away Epistemic Skepticism about Culpability" and "Explaining (Away) the Epistemic Condition on Moral Responsibility."

35 What matters here is simply that these possibilities are relevant alternatives to the state of the actual world at time t . It does not matter whether the actual world could, from an earlier state, evolve into one of these alternative states (given determinism, it of course could not).

possibility that you could instead have had the minimally required quality of will at t (in relation to Y versus Y^*). Similarly, it is a relevant possibility that anyone else involved in the situation who had a poor quality of will at time t could have had the minimally required quality of will at time t . Every combination of these possibilities is relevant. Other possibilities may or may not be relevant as well.³⁶

When we discuss collective harm cases in section 5, we motivate why we have to include the possibility that each agent involved in the situation could have had the required quality of will.³⁷

With this, we may now complete our account of blameworthiness for actions, omissions, and outcomes as follows:

Blameworthiness For: You are blameworthy for X rather than X^* just in case there is a Y and Y^* such that

1. X is worse than X^* , at least partly in virtue of Y being worse than Y^* , and
2. there is a time t such that your poor quality of will at t in relation to Y versus Y^* is a cause of X rather than X^* , within the relevant possibility horizon H .³⁸

We will now test this account on the cases we have considered so far.

Consider first Indifferent John. In this case, (1) the child's drowning *just is* worse than its surviving. Furthermore, (2) John has a poor quality of will at t (the time just before he decides not to intervene) in relation to the child's drowning versus surviving. By Relevant Possibilities for Blame, the relevant possibility

36 This principle presupposes that everyone involved is an agent. If someone is not an agent at the relevant time—if, for instance, he or she is insane or under the influence of drugs—this person may be exempt from the requirement that they should have a particular quality of will. In that case, the relevant possibility horizon for determining blameworthiness may not include the possibility that they could have had a different quality of will. As a result, our account will yield the result that they are not blameworthy. In the following, we set such cases aside.

37 You might wonder how to decide which agents are involved in a situation. Here, it is important to note that Relevant Possibilities for Blame is meant to ensure that enough agents are included. The verdict of our account will not change if even more agents are included. If you are unsure whether an agent is involved, the rule of thumb is to include them.

38 We here set aside cases of deviant causation (for examples, see, e.g., Bernstein, "Moral Luck and Deviant Causation"). To handle such cases, something further is needed—at the very least, a requirement that your bad quality of will is a *nondeviant* cause of X rather than X^* . However, since cases of deviant causation present trouble for everyone, we think it is appropriate to set them aside for now. For discussion, see, e.g., Gunnemyr, *Reasons, Blame, and Collective Harms*, 246–51.

horizon is H_j . And as we have already seen, John's poor quality of will at t in relation to the child's drowning versus surviving is a cause of the child's drowning rather than surviving within H_j . Therefore, Blameworthiness For gives the intuitively correct verdict that John is blameworthy for the death of the child.

Consider next Backup Billy. This case is just like Solo Suzy, except that Billy is lurking in the background. There is a Y and Y^* —the elderly couple's getting upset versus not getting upset—such that (1) the breaking of the window (X) is worse than its staying intact (X^*), at least partly in virtue of the elderly couple's getting upset (Y) being worse than their not getting upset (Y^*). The relevant possibility horizon is H_B , and we have already seen that (2) within H_B , Suzy's poor quality of will at t (the time just before she throws her rock) in relation to the elderly couple's getting upset (Y) versus not (Y^*) is a cause of the window's breaking (X) rather than staying intact (X^*). Thus, Blameworthiness For yields the result that Suzy is blameworthy for the window's breaking rather than staying intact. By contrast, Billy is not blameworthy for this, since there is no process connecting Billy's poor quality of will to the breaking of the window.

Finally, consider Trolley Trouble. Here, (1) the five's getting run over *just is* worse than their not getting run over. Furthermore, (2) Suzy has a poor quality of will at t (the time just before she flips the switch to the left) in relation to the five's getting run over. However, as we have seen, Suzy's poor quality of will at t is *not* a cause of the five's getting run over rather than not within the relevant possibility horizon H_T , since it does not make any difference to the security of their getting run over. Blameworthiness For therefore delivers the intuitively correct result: Suzy is not blameworthy for the five's getting run over rather than not.³⁹

Thus, Blameworthiness For gives the intuitively correct verdict in all three cases.

4. BLAMEWORTHINESS AND FRANKFURT-STYLE CASES

Frankfurt-style cases are important test cases for accounts of blameworthiness.⁴⁰ In these cases, there is a backup ensuring that an agent will act in a

39 This is of course consistent with Suzy's being blameworthy for other things—for example, for intending to kill the five. We may wonder whether it makes any difference for *how* blameworthy Suzy is that, because of circumstances outside of her control, she does not actually kill the five. This is related to the question of moral luck, which we have set aside in this paper.

40 Such cases were first introduced by Frankfurt, "Alternate Possibilities and Moral Responsibility."

certain way. However, as things turn out, the backup does not have to intervene. Consider the following Frankfurt-style variation of Solo Suzy:

Backup Neuroscientist Billy: Everything is as in Solo Suzy, except for the following: unbeknownst to Suzy, the mischievous neuroscientist Billy has implanted a chip in Suzy's brain and is now monitoring her process of deliberation. Billy wants Suzy to throw a rock and break the window. If he thinks, as a result of his monitoring of Suzy's process of deliberation, that Suzy is not going to throw a rock and break the window, he will have the chip induce this intention in her and make her act on it. As things happen, Billy does nothing.

Frankfurt-style cases such as Backup Neuroscientist Billy have a similar structure to early preemption cases such as Backup Billy, where Billy would have thrown a rock at the window had Suzy not done so. In both cases, Billy is lurking in the background with sinister intent. The difference is that, in Backup Neuroscientist Billy, Billy is going to intervene by modifying Suzy's intention, whereas in Backup Billy, he is going to intervene by breaking the window himself. If we consider the process connecting Suzy's poor quality of will to the breaking of the window, we see three salient features. First, there is her poor quality of will, then there is her throw, and finally there is the breaking of the window.

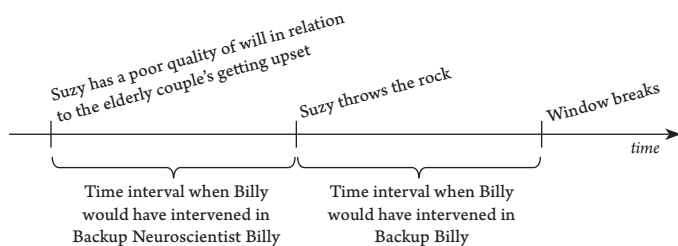


FIGURE 4

Blameworthiness For gives the same result no matter where Billy is ready to enter as a backup in the process connecting Suzy's poor quality of will to the window breaking. This way, Blameworthiness For straightforwardly gives the intuitively right answer in Frankfurt-style cases. In Backup Neuroscientist Billy, for instance, Blameworthiness For gives the verdict that Suzy is blameworthy for breaking the window.

To begin with, (1) the window's breaking is worse than its staying intact, at least partly in virtue of the elderly couple's getting upset being worse than their not getting upset. Furthermore, (2) let t be the time right before Billy would

otherwise have intervened via the chip. At this time, Suzy has a poor quality of will, but it is a relevant possibility that she could have had the required quality of will, and the same goes for Billy. Thus, the relevant possibility horizon is H_{NB} :

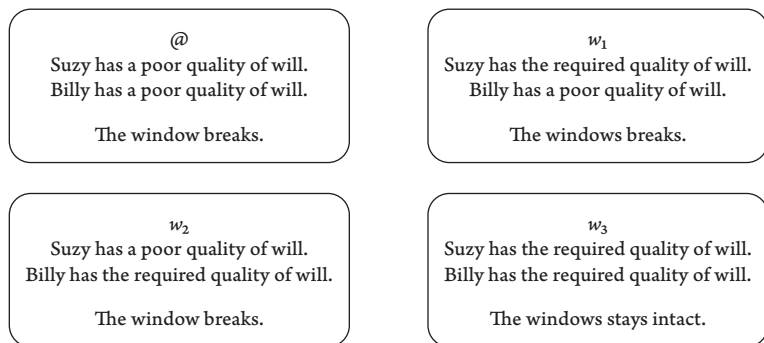


FIGURE 5 Possibility horizon H_{NB}

Applying Causation, we find that Suzy's poor quality of will at t in relation to the elderly couple's getting upset caused the window to break within H_{NB} . We have already seen that (a) Suzy's poor quality of will at t (in relation to the elderly couple's getting upset) is process-connected to the breaking of the window in Backup Billy, and precisely the same reasoning shows that it is also process-connected to the breaking of the window here. Moreover, (b) the window breaking is *less secure* at t in the closest world where Suzy has the minimally required quality of will (namely, w_1): starting from w_1 , only one thing would have to change at t in order for the window not to break (namely, Billy's poor quality of will); starting from @, two things would have to change at t in order for the window not to break (namely, both Suzy's poor quality of will and Billy's poor quality of will). For similar reasons, the window's staying intact is *more secure* at t in the closest possible world where Suzy has the required quality of will. Once again, this reasoning is entirely parallel to the reasoning we applied in the case of Backup Billy.

5. BLAMEWORTHINESS IN COLLECTIVE HARM CASES

Finally, our account sheds light on collective harm cases, i.e., cases where no single action is necessary or sufficient for bringing about some bad outcome. Consider the following case:

The Lake: Ann, Beth, and Claire live close to a lake with a sensitive ecosystem. Each of them has a boat. They have all been using a cheap and hazardous paint. However, they have recently learned that this has brought

the ecosystem to the verge of collapsing. Each of them now believes that if she were to switch to an environmentally friendly but more expensive paint, this might be enough to save the ecosystem from collapse. Still, none of the three cares sufficiently about the ecosystem in the lake. All three continue to use the hazardous paint, and the lake becomes a wet wasteland. As a matter of fact, the ecosystem would still have collapsed if just one of them had switched to the environmentally friendly paint. However, if two or more of them had switched to the environmentally friendly paint, the ecosystem would not have collapsed.⁴¹

Intuitions about this case might vary. It might seem that Ann, Beth, and Claire are blameworthy for the collapse of the ecosystem since the ecosystem collapsed as a result of what they did. It might also seem that none of them is blameworthy since it is true of each that the ecosystem would have collapsed whether or not she had switched to the environmentally friendly paint. If someone, for instance, blames Ann for the collapse of the ecosystem, she might make the following defense:

Defense: "I accept that I may be blameworthy for using the hazardous paint. But I am not blameworthy for the collapse of the ecosystem. Given that Beth and Claire did not care, my lack of care did not matter."⁴²

Moreover, even if we think that Ann, Beth, and Claire are blameworthy for the collapse of the ecosystem, there is a puzzle about whether they are to blame individually or collectively. On the one hand, it seems that they are not individually to blame since each can appeal to the defense just sketched. On the other hand, it seems that they could not be collectively to blame since they do not constitute a collective. They did not perform an intentional collective action, and they do not share a formal decision procedure.⁴³

41 Adapted from Björnsson, "Joint Responsibility Without Individual Control."

42 Such a defense was considered in an early version of Björnsson, "Being Implicated," presented at the Manchester Centre for Political Theory (MANCEPT), 2019. Björnsson accepts the defense insofar as it concerns explanation: applied to The Lake, his view in "Being Implicated" is that we should reject the claim that "the ecosystem collapsed *because* Ann did not care." However, he argues that Ann is still a fitting target of reactive attitudes over the outcome because her substandard care is *involved* in a normal explanation of the ecosystem's collapse.

43 According to standard accounts of intentional collective action, they did not perform a collective action. See Bratman, *Shared Agency*; Gilbert, *Joint Commitment*; and Searle, "Collective Intentions and Actions." For accounts of what it is to share a formal decision procedure, see French, *Collective and Corporate Responsibility*; and List and Pettit, *Group Agency*.

We suggest that the conflicting intuitions about The Lake can be understood as arising from different choices of possibility horizon—where the choice of possibility horizon may, in turn, reflect either a collective or an individual perspective. Consider the time t before any of the three has painted their boats. Following the standard procedure for generating the relevant possibility horizon based on what is happening at time t (as stated in Relevant Possibilities for Blame), we get a possibility horizon that contains every combination of Ann, Beth, and Claire having their actual poor quality of will at t , and having the minimally required quality of will at t . Call this possibility horizon H_{large} . In some worlds within this possibility horizon, the ecosystem will be saved. In other worlds, the ecosystem will collapse (as in the actual world).

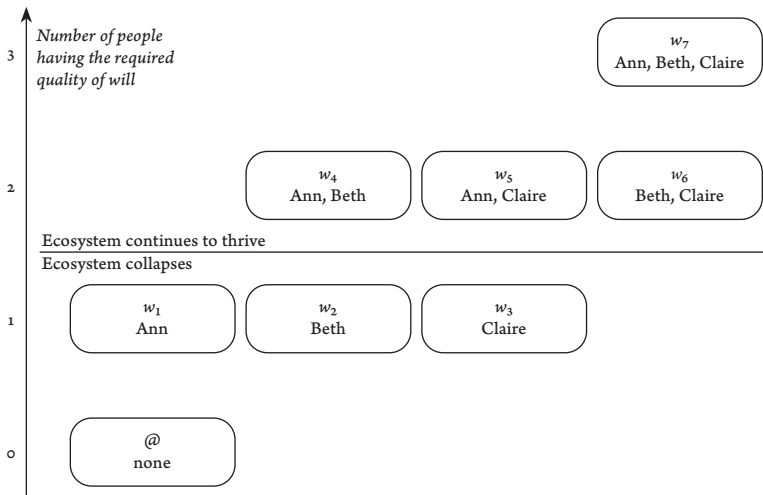


FIGURE 6 Possibility horizon H_{large}

Now consider Ann. When we use H_{large} , we find that Ann is blameworthy for the collapse of the ecosystem, since (1) the collapse of the ecosystem *just is* worse than its continuing to thrive, and (2) Ann’s poor quality of will at t in relation to the ecosystem’s collapsing versus continuing to thrive caused the ecosystem to collapse rather than thrive. In support of 2, note that (a) Ann’s poor quality of will at time t is process-connected to the collapse of the ecosystem: her poor quality of will is, together with, for instance, Beth’s poor quality of will, minimally sufficient for the collapse of the ecosystem (likewise, Ann’s and Claire’s poor qualities of will are minimally sufficient for the collapse), and this connection remains when we consider more intermediate times. Furthermore, (b) the collapse of the ecosystem is *less secure* at t in the closest world where Ann does not have a poor quality of will, i.e., in world w_1 , where she has the required

quality of will. Here, only one instead of both of the other boat owners would also need to have the required quality of will in order for the ecosystem not to collapse. For similar reasons, the ecosystem's continuing to thrive is *more secure* at t in the closest world where Ann has the required quality of will. Thus, Blame-worthiness For yields the result that Ann is blameworthy for the ecosystem's collapsing rather than continuing to thrive. The same applies to Beth and Claire.

However, Ann might argue that H_{large} is not the relevant possibility horizon. As we saw above, she might defend herself by saying something like the following: "given that Beth and Claire did not care . . ." There is a straightforward reading of this statement in terms of which worlds should be included in the relevant possibility horizon: Ann is saying, essentially, that we should hold fixed that the others did not care about the lake and treat this as a background condition. If we comply with this request, we get a much smaller possibility horizon, H_{small} . This possibility horizon contains just two possible worlds: the actual world and a world where Ann has the minimally required quality of will. The ecosystem collapses in both worlds since Beth and Claire continue to use the hazardous paint.

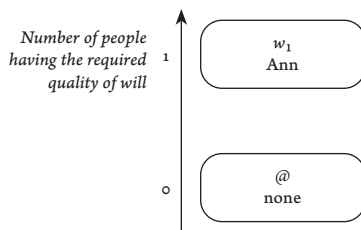


FIGURE 7 Possibility horizon H_{small}

Within H_{small} , the collapse of the ecosystem is *just as secure* at t in the actual world as it is in the closest world where Ann has the required quality of will—namely, infinitely secure. Thus, condition b of Causation is not satisfied. Given H_{small} , which holds fixed the motivations of the other boat owners, Blame-worthiness For therefore gives the verdict that Ann is *not* blameworthy for the collapse of the ecosystem.

This suggests that our conflicting intuitions about The Lake can be explained as the result of employing two different possibility horizons: employing the large possibility horizon H_{large} yields the result that Ann (as well as Beth and Claire) is individually blameworthy for the collapse of the ecosystem. Employing the small possibility horizon H_{small} , which treats Ann's quality of will as variable while holding the quality of will of the other boat owners fixed, yields the result that Ann is *not* individually blameworthy for the collapse of the

ecosystem. Thus, it matters to our assessment of Ann's *individual* responsibility whether we consider a possibility horizon that treats her as part of a group where each member's poor quality of will is a candidate cause, or whether we instead consider a possibility horizon that treats her poor quality of will as the *only* candidate cause. The conflicting intuitions are explained, not as arising from two kinds of entities that can be blameworthy—the group and the individual—but instead as arising from two different perspectives we can take when we are assessing the blameworthiness of an individual.

Of course, we cannot say that Ann both is and is not individually blameworthy for the collapse of the ecosystem. We need to single out either H_{large} or H_{small} as being the relevant possibility horizon for assessing Ann's blameworthiness. We have already suggested in *Relevant Possibilities for Blame* that we should treat it as a relevant possibility that everyone involved could have had the required quality of will. If this is accurate, the larger possibility horizon is the correct one. We will now support this with two considerations.

First, given that it is important to understand what brought about the collapse of the ecosystem, we have reasons to widen and adjust H_{small} , while H_{large} can do the job satisfactorily. If we address Ann alone, her defense that her lack of care did not matter seems persuasive. At least, it seems to be an open question whether we should consider the smaller possibility horizon she insists on, or the larger one. However, Beth could also argue in the same way that *she* is not blameworthy for the collapse of the ecosystem. And so could Claire. The situation is symmetrical, so if the defense is open to one, it is open to all. If we accept Ann's defense and follow this argument to its logical conclusion, we end up concluding that neither Ann, nor Beth, nor Claire is a cause of the collapse of the ecosystem, and therefore none of them is blameworthy. At this point, we might suspect that something has gone wrong. Surely, we might think, some of the boat owners must be blameworthy for the collapse of the ecosystem. After all, it seems clear that *their* lack of care for the ecosystem played a crucial role in bringing about its collapse.⁴⁴ So we face two (related) puzzles: one concerning causation and one concerning blameworthiness. When facing such puzzles, we think there are reasons to reconsider one's choice of possibility horizon:

Widening and Adjusting: If it is important to understand what brought about X , and if the causes of X are unsatisfactorily explained, we have reasons to look for more possibilities to include in the possibility horizon under consideration and to scrutinize the possibilities we have already included.

44 See Björnsson, "Joint Responsibility Without Individual Control," 192–93.

If we accept Ann's defense, and the similar defenses of the other boat owners, we find no cause of the collapse and no one who is blameworthy. If it is important to understand what brought about the collapse of the ecosystem, this gives us reason to turn to the larger possibility horizon H_{large} . Within H_{large} , we do find a satisfactory explanation—namely, that all three boat owners are causes of and blameworthy for the collapse of the ecosystem.

Importantly, the reason to widen and adjust our possibility horizon is not grounded in the fact that widening the possibility horizon allows us to hold Ann, Beth, and Claire responsible for the collapse, or say that they caused the collapse. If so, the reasoning would be circular. Rather, the reason to widen and adjust our possibility horizon is grounded in the fact that it is important to find out what brought about the collapse of the ecosystem. Maybe we seek to ensure that similar things do not happen in the future, or maybe we seek to find out who, if anyone, is to blame for the collapse of the ecosystem. If it would turn out upon closer scrutiny—even after widening the possibility horizon—that only Ann caused the collapse, or that none of them did (perhaps something else entirely caused the collapse), this is the conclusion we should accept.

Second, the smaller possibility horizons that Ann, Beth, and Claire must appeal to in their defenses do not fit together. As stated before, if we address Ann alone, her defense seems persuasive. However, this defense is only open to Ann as long as we treat Beth and Claire's poor quality of will as fixed. Ann's defense presupposes that *her* poor quality of will is a candidate cause, while Beth's (and Claire's) poor quality of will is a mere background condition. Beth's defense, on the other hand, presupposes that *her* poor quality of will is a candidate cause, while Ann's (and Claire's) poor quality of will is a mere background condition. In order to accept Ann, Beth, and Claire's defenses, we thus have to adopt one perspective when evaluating whether Ann is blameworthy, a different perspective when evaluating whether Beth is blameworthy, and a third when evaluating Claire's blameworthiness. But these perspectives do not fit together. We cannot at the same time treat, e.g., Ann's quality of will as a candidate cause and as a mere background condition.⁴⁵ Given the social function of blame, it should be possible to assess blameworthiness from a single perspective that everyone involved in the situation can accept, and that can be used to assess the blameworthiness of everyone involved. This disqualifies H_{small} since it does not provide such a perspective.

We should point out that there would be no problem of possibility horizons not fitting together if we were to consider a possibility horizon where

45 Jamieson proposes a similar argument in "When Utilitarians Should Be Virtue Theorists," 176.

the poor quality of will of *each* of Ann, Beth, and Claire was treated as a mere background condition. Thus, there are in fact two ways to treat Ann, Beth, and Claire equally: treating them all as candidate causes (as in H_{large}), or treating them all as mere background conditions. If we, for instance, were to evaluate whether the company producing the hazardous paint is blameworthy for the collapse of the ecosystem, we might for the sake of simplicity treat the poor quality of will of Ann, Beth, and Claire as a mere background condition and instead focus on the quality of will of the company. However, when the question at issue is whether, e.g., Ann is to blame for the collapse of the ecosystem, there is a compelling reason not to treat Ann's poor quality of will as a mere background condition: doing so would *prejudge* the question of whether she is blameworthy by not treating her poor quality of will as a candidate cause at all. Furthermore, we would be failing to treat Ann as an agent if we were to reject the possibility that she could have had the required quality of will. Thus, when Ann's blameworthiness is at issue, we have to treat Ann's poor quality of will as a candidate cause; and then we also have to treat Beth's poor quality of will and Claire's poor quality of will as candidate causes.

We therefore conclude that H_{large} is the relevant possibility horizon. Thus, Blameworthiness For entails that Ann is directly individually blameworthy for the collapse of the ecosystem. By contrast, writers like Held and Wringer would argue that Ann is only *indirectly* individually blameworthy for this outcome.⁴⁶ On their view, individual blameworthiness in cases like *The Lake* derives from the blameworthiness of the group. This presupposes that unstructured groups—i.e., groups that lack collective intentions or formal decision procedures—also can be blameworthy. Blameworthiness For does not presuppose this controversial idea. Still, our view captures Held's and Wringer's crucial insight that individual blameworthiness disappears from the scene if we lose sight of the group. For instance, if we fail to consider the possibility that each of Ann, Beth, and Claire could have cared enough about life in the lake, we may end up basing our assessment of Ann's blameworthiness on H_{small} , according to which Ann is not blameworthy for the collapse of the ecosystem.

6. CONCLUSION

We have suggested that Blameworthiness For together with Causation gives an accurate account of when you are blameworthy for an action, omission, or

46 See Held, "Can a Random Collection of Individuals Be Morally Responsible?"; and Wringer, "Collective Obligations."

outcome. This account captures the intuitive idea that you are blameworthy for something if it happened because you did not care enough.

One virtue of the account is that it gives the right verdict in a wide range of cases. These include cases of forgetting, making a negative difference to a nevertheless good result (as when Sally did not pay attention to the chili recipe), or doing an action with runaway consequences (like Tragedy). They also include cases where we disregard irrelevant possibilities (like when we think the queen of Sweden did not cause Selma's flowers to die), omission cases (like Indifferent John), switching cases (like Trolley Trouble), (early) preemption cases (like Backup Billy), Frankfurt-style cases (like Backup Neuroscientist Billy), and collective harm cases (like The Lake). In addition, our account gives the right verdict in a wide range of other cases which we do not have space to discuss here, including cases of overdetermination, late preemption, and double prevention, as well as collective harm cases without a threshold.⁴⁷

Another virtue of the account is that it can explain the conflicting intuitions we have about some cases. In collective harm cases like The Lake, for instance, it explains why it might be tempting to accept Ann's defense that we should not blame her: if we treat it as a background condition that Beth and Claire did not care, as Ann insists we should, Ann is correct that *her* lack of care did not cause the collapse. At the same time, our account also explains why it seems that Ann, Beth, and Claire *are* individually blameworthy for the collapse: when we treat it as a relevant possibility that each of them could have had the required quality of will, we find that each of them is a cause of the bad outcome.

As might be evident by now, our account entails that it matters for blameworthiness which possibilities are relevant. Correspondingly, it matters for our judgments about blameworthiness which possibilities we consider to be relevant. This explains, for example, why our intuitions are torn in The Lake: we are torn between only treating it as relevant that Ann could have had the required quality of will, or treating it as relevant that each of the three could have had the required quality of will. To say something about blameworthiness itself, rather than merely about our judgments, we need to answer the question: Which possibilities are relevant? We have argued that whenever an agent involved in the situation has a poor quality of will, it is a relevant possibility that they could instead have had the required quality of will, and the relevant possibility horizon includes, as a minimum, worlds representing every combination of

47 For a discussion of such cases, see Gunnemyr, *Reasons, Blame, and Collective Harms*, chs. 12–13.

such possibilities. This means, for example, that the relevant possibility horizon in The Lake is the larger one, and so Ann, Beth, and Claire are all individually blameworthy for the collapse of the ecosystem.⁴⁸

Lund University
gunnemyr@gmail.com

Umeå University
caroline.touborg@gmail.com

REFERENCES

- Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press, 2002.
- Bernstein, Sara. "Moral Luck and Deviant Causation." *Midwest Studies in Philosophy* 43, no. 1 (September 2019): 151–61.
- Björnsson, Gunnar. "Being Implicated: On the Fittingness of Guilt and Indignation over Outcomes." *Philosophical Studies* 178, no. 11 (November 2021): 3543–60.
- . "Explaining Away Epistemic Skepticism about Culpability." In *Oxford Studies in Agency and Responsibility*, edited by David Shoemaker, 141–64. Oxford: Oxford University Press, 2017.
- . "Explaining (Away) the Epistemic Condition on Moral Responsibility." In *Responsibility: The Epistemic Condition*, edited by Philip Robichaud and Jan Willem Wieland, 146–62. Oxford: Oxford University Press, 2017.
- . "Joint Responsibility Without Individual Control: Applying the Explanation Hypothesis." In *Moral Responsibility: Beyond Free Will and*

⁴⁸ We would like to thank Gunnar Björnsson, Erik Carlson, Anton Emilsson, Jens Johansson, and Per-Erik Milam for their close reading and valuable comments, and we are grateful to our colleagues at the Lund Gothenburg Responsibility Project. We are also grateful to audiences at the Group Agency and Collective Responsibility workshop in Flensburg, the Group Agency and Collective Responsibility workshop in Gothenburg, the Higher Seminar in Practical Philosophy at Lund University, the conference Social Ontology 2020, and the Higher Seminar in Practical Philosophy at Uppsala University. We would especially like to thank Per Algander, Franz Altner, Emil Andersson, Olle Blomberg, Dan Egonsson, Anna Folland, Frits Gåvertsson, Niels de Haan, Giulia Lasagni, Samuel Lee, Marianna Leventi, Anne-Sofie Munk Autzen, Carlos Núñez, Anna Nyman, Grace Paterson, Andrew Peet, Björn Petersson, Matthew Rachar, Wlodek Rabinowicz, Toni Rønnow-Rasmussen, Andrew Reisner, David Schweikard, Matthew Talbert, Alexander Velichkov, Marta Johansson Werkmäster, Jakob Werkmäster, and Michael Wilby. Finally, we would like to thank an anonymous reviewer from this journal.

- Determinism*, edited by Jeroen van den Hoven, Ibo van de Poel, and Nicole Vincent, 181–99. Dordrecht: Springer, 2011.
- Bratman, Michael. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press, 2014.
- Dowe, Phil. *Physical Causation*. Cambridge: Cambridge University Press, 2000.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998.
- Foot, Philippa. “The Problem of Abortion and the Doctrine of Double Effect.” *Oxford Review* 5 (1967): 5–15.
- Frankfurt, Harry G. “Alternate Possibilities and Moral Responsibility.” *Journal of Philosophy* 66, no. 23 (December 1969): 829–39.
- French, Peter. *Collective and Corporate Responsibility*. New York: Columbia University Press, 1984.
- Gilbert, Margaret. *Joint Commitment: How We Make the Social World*. Oxford: Oxford University Press, 2014.
- Gunnemyr, Mattias. *Reasons, Blame, and Collective Harms*. PhD Thesis, Lund University, 2021. <https://portal.research.lu.se/en/publications/reasons-blame-and-collective-harms>.
- Hall, Ned. “Structural Equations and Causation.” *Philosophical Studies* 132, no. 1 (January 2007): 109–36.
- . “Two Concepts of Causation.” In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 225–76. Cambridge, MA: MIT Press, 2004.
- Hart, H. L. A., and Tony Honoré. *Causation in the Law*. Oxford: Clarendon Press, 1985.
- Held, Virginia. “Can a Random Collection of Individuals Be Morally Responsible?” *Journal of Philosophy* 67, no. 14 (July 1970): 471–81.
- Jamieson, Dale. “When Utilitarians Should Be Virtue Theorists.” *Utilitas* 19, no. 2 (June 2007): 160–83.
- Lewis, David. “Causation.” *Journal of Philosophy* 70, no. 17 (October 1973): 556–67.
- . “Causation as Influence.” In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 75–106. Cambridge, MA: MIT Press, 2004.
- List, Christian, and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press, 2011.
- McGrath, Sarah. “Causation by Omission: A Dilemma.” *Philosophical Studies* 123, nos. 1–2 (March 2005): 125–48.
- McKenna, Michael. *Conversation and Responsibility*. Oxford: Oxford University Press, 2012.

- Paul, L. A., and Ned Hall. *Causation: A User's Guide*. Oxford: Oxford University Press, 2013.
- Rosen, Gideon. "The Alethic Conception of Moral Responsibility." In *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 65–87. Oxford: Oxford University Press, 2015.
- Sartorio, Carolina. *Causation and Free Will*. Oxford: Oxford University Press, 2016.
- . "Causes as Difference-Makers." *Philosophical Studies* 123, nos. 1–2 (March 2005): 71–96.
- Schaffer, Jonathan. "Causal Contextualism." In *Contrastivism in Philosophy*, edited by Martijn Blaauw, 35–63. Abingdon: Routledge, 2013.
- . "Causation by Disconnection." *Philosophy of Science* 67, no. 2 (June 2000): 285–300.
- . "Contrastive Causation." *Philosophical Review* 114, no. 3 (July 2005): 327–58.
- Searle, John. "Collective Intentions and Actions." In *Intentions in Communication*, edited by Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, 401–15. Cambridge, MA: MIT Press, 1990.
- Shoemaker, David. *Responsibility from the Margins*. Oxford: Oxford University Press, 2015.
- Smith, Angela M. "Attitudes, Tracing, and Control." *Journal of Applied Philosophy* 32, no. 2 (May 2015): 115–32.
- . "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (January 2005): 236–71.
- Strawson, P. F. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1–28. Abingdon, UK: Routledge, 2008.
- Talbert, Matthew. "Moral Competence, Moral Blame, and Protest." *Journal of Ethics* 16, no. 1 (March 2012): 89–109.
- Thomson, Judith Jarvis. "Killing, Letting Die, and the Trolley Problem." *Monist* 59, no. 2 (April 1976): 204–17.
- Touborg, Caroline Torpe. *The Dual Nature of Causation: Two Necessary and Jointly Sufficient Conditions*. PhD Thesis, University of St Andrews, 2018. <http://hdl.handle.net/10023/16561>.
- . "Hasteners and Delayers: Why Rains Don't Cause Fires." *Philosophical Studies* 175, no. 7 (July 2018): 1557–76.
- Van Inwagen, Peter. "Ability and Responsibility." *Philosophical Review* 87, no. 2 (April 1978): 201–24.
- Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.

- Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand Schoeman, 256–86. Cambridge: Cambridge University Press, 1988.
- Woollard, Fiona. *Doing and Allowing Harm*. Oxford: Oxford University Press, 2015.
- Wringe, Bill. "Collective Obligations: Their Existence, Their Explanatory Power, and Their Supervenience on the Obligations of Individuals." *European Journal of Philosophy* 24, no. 2 (June 2016): 472–97.
- Zimmerman, Michael J. "Taking Luck Seriously." *Journal of Philosophy* 99, no. 11 (November 2002): 553–76.