Erkenntnis
https://doi.org/10.1007/s10670-023-00660-z

ORIGINAL RESEARCH

Actual Causation and the Challenge of Purpose

Enno Fischer[1]

Received: 30 May 2022 / Accepted: 7 January 2023

Abstract
This paper explores the prospects of employing a functional approach in order to improve our concept of actual causation. Claims of actual causation play an important role for a variety of purposes. In particular, they are relevant for identifying suitable targets for intervention, and they are relevant for our practices of ascribing responsibility. I argue that this gives rise to the *challenge of purpose*. The challenge of purpose arises when different goals demand adjustments of the concept that pull in opposing directions. More specifically, I argue that a common distinction between certain kinds of preempted and preempting factors is difficult to motivate from an interventionist viewpoint. This indicates that an appropriately revised concept of actual causation would not distinguish between these two kinds of factors. From the viewpoint of retributivist responsibility, however, the distinction between preempted and preempting factors sometimes is important, which indicates that the distinction should be retained.

1 Introduction

At the intersection of philosophy, law, computer science, and empirical psychology there is a growing literature aiming to establish a definition of actual causation that adequately captures our—often linguistic—causal intuitions. The formal framework of causal models has without doubt advanced this enterprise, yet, no consensus has emerged. This has given rise to a methodological shift towards functional approaches. Functional approaches do not aim at a mere description of our causal intuitions but typically take an evaluative stance. Given that the concept of actual causation fulfils such and such a role in the cognitive life of causal reasoners, what features *should* it have? James Woodward (2021), for example, suggests that we can view causal concepts as a kind of epistemic technology or tool, and that we can evaluate and understand causal concepts in terms of how well they facilitate our goals.

✉ Enno Fischer
enno.fischer@rub.de

[1] Institut für Philosophie I, Ruhr-Universität Bochum, Universitätsstr. 150, Bochum, Germany

Published online: 01 February 2023

In this paper I will explore the prospects of employing a functional approach in order to *improve*[1] our concept of actual causation. Taking up Woodward's 'tool' metaphor, I will investigate whether and under which circumstances the concept of actual causation can be engineered such that its performance is improved. I will thus connect the functional project to a recent revival of interest in Carnapian explication and discussions of conceptual engineering. I will argue that functional approaches to actual causation face what I call the *challenge of purpose*. There is not a uniquely correct way to improve the concept of actual causation but many ways that can be more or less fruitful. Fruitfulness, in turn, depends on the particular context and goal of application. The challenge of purpose arises when different goals demand adjustments that pull in opposing directions.

The concept of actual causation has important roles to play for a variety of goals. We employ the concept, for example, in order to highlight suitable targets for intervention but also in order to support our evaluations of responsibility. I will first look at the concept of actual causation from the viewpoint of intervention. This discussion will suggest a potential revision of the concept. However, if we tune the concept of actual causation such that its performance with regard to facilitating intervention is improved, then that does not mean that it will also be improved with regard to all other potential goals. In fact, I will argue that there are instances where attempts to improve the concept with regard to intervention may even have adverse consequences for the usefulness of the concept with regard to considerations of responsibility.

Claims of actual causation relate actual events or states of affairs to each other, as in 'the bottle's shattering was caused by the impact of the stone'. In this regard such claims are different from causal laws and regularities commonly studied in the sciences. The concept of actual causation is also commonly thought to be sensitive to considerations of preemption. Suppose Suzy and Billy each throw a stone at a bottle. Suzy's stone hits the bottle and shatters it. Billy throws a split second later and his stone passes the initial position of the bottle. If Suzy's stone had not destroyed the bottle, then it would have been hit and destroyed by Billy's stone (Hall, 2004). In this kind of situation, we typically identify Suzy as actual cause of the bottle's shattering but not Billy. Approaches aiming at a descriptively adequate definition of actual causation have been struggling with this kind of 'late preemption' but by now there are multiple approaches (Pearl, 2000; Woodward, 2003; Halpern & Pearl, 2005; Halpern & Hitchcock, 2015; Halpern, 2016; Beckers & Vennekens, 2018; Beckers, 2021; Andreas & Günther, 2021; Gallow, 2021).

But why do we care about the difference between Suzy and Billy in the first place? Interventionists argue that we generally tend to care about the concept of actual causation because it indicates factors that are particularly suitable as targets for intervention if we aim to alter or prevent a certain outcome (Hitchcock & Knobe, 2009; Hitchcock, 2017). However, I will argue that the interventionist view does not help to explain the perceived difference between Suzy and Billy. If one aims to prevent the bottle's

[1] The idea of improving our concept of actual causation assumes that we have a concept of actual causation, but that this concept may not be the best possible concept. Alternatively, one can construe the functional project as aiming at *discovering* the best possible concept of actual causation.

shattering, Suzy is not a better target of intervention than Billy. Instead, one would have to intervene on both Suzy *and* Billy. I will take this to indicate that on the interventionist view there are good reasons to prefer a concept that would not rule out a preempted factor like Billy.

However, we will see that this conclusion is problematic for reasons related to the challenge of purpose. More specifically, there is a natural understanding of the situation on which there *is* an important difference between Suzy and Billy, because only Suzy is responsible for the bottle's shattering. That is, from the viewpoint of responsibility there appear to be good reasons to prefer a causal concept that does rule out a preempted factor like Billy. A potential objection here is that Billy is just as blameworthy as Suzy because he attempted to destroy the bottle. We will see that this objection may reflect a certain consequentialist conception of responsibility. However, there are also approaches to responsibility according to which the difference between preempted and preempting factors is relevant. More specifically, I will look at theories of retributive justice, that is, theories that ascribe an intrinsic moral value to holding wrongdoers responsible. Thus, whether the difference is important, depends on one's understanding of responsibility.

What are the consequences of the challenge of purpose? I do not think that the challenge poses an insurmountable problem to functional approaches. But the challenge of purpose indicates that functional approaches are unlikely to provide a one-size-fits-all concept of actual causation. Instead, the challenge of purpose may indicate the need for some kind of pluralist theory of actual causation, that is, a theory of causation according to which we should have more than one concept of actual causation and that these concepts serve different purposes.

In Sect. 2 I will briefly introduce the formal framework of causal models and provide an illustrative definition of actual causation, as proposed by Halpern and Hitchcock (2015). In Sect. 3 I will suggest that the challenge of purpose arises for functional approaches if they are pursued with the aim of revising the concept of actual causation. In Sect. 4 I will evaluate the definition presented in Sect. 2 from the viewpoint of an intervening agent. Looking at cases of late preemption I will argue that the definition's distinction between preempted and preempting factors is difficult to motivate from this viewpoint. I will also argue that other popular definitions of actual causation face similar problems. In Sect. 5 I will take the viewpoint of responsibility. From this perspective the distinction between preempted and preempting factors is important, at least on a retributivist understanding of responsibility. I conclude in Sect. 6 that functional approaches to actual causation that are pursued with the goal to improve the concept may come to different results, depending on whether they are oriented towards facilitating intervention or towards facilitating judgements of responsibility. As an outlook I suggest that a potential solution of the challenge of purpose is to pursue a pluralist account of actual causation.

Fig. 1 A causal model of late preemption

2 Actual Causation

A central project in the philosophy of causation has been to find necessary and sufficient conditions under which some event c is an actual cause of another event e. Since David Lewis's (1973) seminal account, a dominant approach is to spell out these conditions in terms of counterfactual conditionals. One central challenge to this approach arises from instances that involve redundancy. Consider the example given in the introduction. We typically say that Suzy's throwing her stone is the actual cause of the bottle's shattering. This is the case even though the bottle's shattering does not depend counterfactually on Suzy's actions. If she had not hit the bottle, then Billy would have hit it and destroyed it.

Many contemporary definitions of actual causation have been proposed within the formal framework of causal models (Spirtes et al., 1993; Pearl, 2000). Formally, a causal model consists of a set of variables and a set of structural equations that summarize the dependencies between the variables. The variables are represented by upper case letters (X, Y, $\cdots$) and typically take on one of two or more possible values (lower case letters x, y, $\cdots$) that represent possible events or states of affairs in the model's target system. We distinguish between a set of exogenous variables $\mathcal{U}$ and a set of endogenous variables $\mathcal{V}$. The values of the exogenous variables are determined by factors outside the model, and the values of the endogenous variables are determined by the exogenous variables or other endogenous variables.

Figure 1 shows a graph and the structural equations of the example of late preemption. All variables have two possible values, describing whether the corresponding event does occur (1) or does not occur (0). In particular, ST and BT represent whether Suzy and Billy throw their stones, respectively; and SH and BH describe whether Suzy's and Billy's stones hit the bottle, respectively. Finally, variable BS represents whether the bottle is shattered or not. If Suzy hits the bottle, the bottle shatters ($BS = 1$). As a result, Billy's stone will not hit the bottle ($BH = 0$). If Suzy does not hit the bottle, then it will be hit by Billy's stone and still be destroyed ($BS = 1$).[2]

As a basis for the following discussion we shall employ a definition of actual causation proposed by Halpern and Hitchcock (2015), which defines an actual cause of

[2] Note that this model slightly misrepresents the case. It does not capture the fact that Billy's not hitting the bottle is caused by the bottle's already being destroyed when Billy's stone arrives.

some event φ relative to a model M and a value assignement $\vec{u}$ to the model's exogenous variables as follows.[3]

$\vec{X} = \vec{x}$ is an actual cause of φ in $(M, \vec{u})$ if the following three conditions hold:

AC1: $(M, \vec{u}) \models (\vec{X} = \vec{x}) \wedge \varphi$

AC2: There exists a partition $(\vec{Z}, \vec{W})$ of $\mathcal{V}$ with $\vec{X} \subseteq \vec{Z}$ and some setting $(x', \vec{w}')$ of the variables in $(\vec{X}, \vec{W})$ such that if $(M, \vec{u}) \models Z = z^*$ for all $Z \in \vec{Z}$, then both of the following conditions hold:

(a) $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}']\neg\varphi$ and $s_{\vec{X}=\vec{x}', \vec{W}=\vec{w}', \vec{u}} \succeq s_{\vec{u}}$.

(b) $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W}' \leftarrow \vec{w}', \vec{Z}' \leftarrow \vec{z}^*]\varphi$ for all subsets $\vec{W}'$ of $\vec{W}$ and all subsets $\vec{Z}'$ of $\vec{Z}$.[4]

AC3: $\vec{X}$ is minimal. No subset of $\vec{X}$ satisfies conditions AC1 and AC2.

Condition AC1 states that in the situation under consideration (described by model M and the values of the model's exogenous variables $\vec{u}$) both the actual cause $\vec{X} = \vec{x}$ and the effect event φ need to occur.[5] For example, Suzy's throwing her stone can only be an actual cause of the bottle's shattering if in the model's target system Suzy actually throws her stone ($ST = 1$) and the bottle is actually shattered ($BS = 1$).

Condition AC2(a) is a requirement that is more permissive than direct counterfactual dependence: we may change the values of other variables $\vec{W}$, in order to unmask a counterfactual dependence of the outcome φ on the actual cause $\vec{X}$. For example, we can imagine a situation in which Billy does not hit the bottle, independently of Suzy's actions (that is, we can choose $\vec{W} = \vec{w}'$ such that $BH = 0$ is held fixed). In this situation the outcome φ *does* depend upon Suzy's action, which means that Suzy's actions satisfy condition AC2(a) (ignore for a moment the additional requirement that $s_{\vec{X}=\vec{x}', \vec{W}=\vec{w}', \vec{u}} \succeq s_{\vec{u}}$, which will be explained below). On its own, condition AC2(a) would be too permissive, however, because it would also identify Billy as actual cause. We can imagine a situation in which Suzy does not throw her stone (that is, we can choose $\vec{W} = \vec{w}'$ such that $ST = 0$). In this situation the outcome depends

[3] The definition is closely related to a definition proposed by Halpern and Pearl (2005), also known as the HP definition of actual causation. Other definitions have been proposed by Pearl (2000); Hitchcock (2001); Woodward (2003); Hitchcock (2007a); Halpern (2016); Beckers and Vennekens (2018); Andreas and Günther (2021); Gallow (2021). There is an ongoing debate about which definition best captures our causal intuitions. My reason for choosing Halpern and Hitchcock's (2015) definition here is that their distinction between AC2(a) and AC2(b) makes it particularly straightforward to develop the following argument, because our main concern will be the intuitions captured by AC2(b). Towards the end of Sect. 4 I will discuss how the results of my approach affect definitions of actual causation that do not involve condition AC2(b).

[4] As Halpern and Hitchcock (2015) point out, notation is slightly abused here: $\vec{W}' \leftarrow \vec{w}'$ denotes the assignment where the variables in $\vec{W}'$ get the same values as they would in the assignment $\vec{W} \leftarrow \vec{w}'$, and similarly for $\vec{Z}$.

[5] The vector notation means that the cause event may be an assignment of values to a set of variables $X_i \in \vec{X}$.

upon Billy's actions, even though we would not say that Billy is an actual cause in the original situation.

This is why we need the further restriction introduced by AC2(b), which takes into consideration the variables $\vec{Z}$ that lie on the path linking the cause $\vec{X} = \vec{x}$ and the outcome φ. AC2(b) requires that the outcome φ be sustained if we set $\vec{W}' = \vec{w}'$ while the cause variable $\vec{X}$ and any subset of $\vec{Z}$ is held fixed at its original values ($\vec{x}$ and $\vec{z}^*$, respectively). The rationale is that the causal process initiated by the actual cause should be sufficient to sustain the effect even if we apply the variation of $\vec{W}$ (and any possible subset of it) that is required to unmask the counterfactual dependence.[6]

Finally, Halpern and Hitchcock (2015) propose including a further restriction to condition AC2(a) which requires that the counterfactual scenario considered in AC2(a) (denoted by $s_{\vec{X}=\vec{x}',\vec{W}=\vec{w}'}$, and called the 'witness') be at least as normal as the actual scenario (denoted by $s_{\vec{u}}$). This additional restriction (henceforth called the normality criterion) excludes counterfactual scenarios that involve remote possibilities. Here is a standard example: suppose there is a fire that would not have occurred if either a short circuit had not occurred or if sufficient oxygen had not been present. Both the short circuit and the oxygen are necessary conditions, but we typically identify the short circuit as actual cause and the presence of oxygen as a mere background condition. The normality criterion captures this intuition. The 'witness' of the short circuit is a situation where no fire occurs because there is no short circuit. If we suppose that short circuits are exceptional, then this amended situation is at least as normal as the actual scenario. The 'witness' of oxygen is a situation where no fire occurs because there is no oxygen. But the absence of oxygen is (at least in most circumstances) exceptional, which means that this amended situation is less normal than the actual world.

Note that Halpern and Hitchcock here employ a concept of normality that is deliberately wide. It comprises a descriptive dimension—in the sense of statistical frequency—but also a range of normative dimensions in the sense of functional, legal, social, and moral norms. Several questions arise here. What happens in cases where these different dimensions of normality stand in conflict? Should the concept of actual causation be related to considerations of normality at all? Doesn't this make the concept unduly subjective and context sensitive?

So far, the discussion has been concerned with the descriptive project. This project takes intuitions such as those regarding Billy and Suzy as basic data that are to be captured in terms of necessary and sufficient conditions for the correct use of the concept 'actual cause'. The descriptive project has certainly advanced our understanding of what an actual cause is, but it also faces a number of problems (Glymour

[6] More concretely, Billy is not an actual cause because on any choice of the partition $(\vec{Z}, \vec{W})$ that fulfils condition AC2(a) it will be the case that condition AC2(b) is violated. The only admissible choice of the partition $(\vec{Z}, \vec{W})$ from the viewpoint of AC2(a) is $\vec{W} = \{ST, SH\}$ and $\vec{Z} = \{BT, BH, BS\}$. In order to have a counterfactual dependence of BS on BT, we need to set $\vec{w}'$ such that $ST = 0$, $SH = 0$. But then condition AC2(b) is not fulfilled. There is a subset $\vec{Z}' = \{BH\}$ of $\vec{Z}$ such that if we fix $\vec{Z}' = \vec{z}^*$ (that is: $BH = 0$), then the bottle will not be shattered ($BS = 0$). Suzy's throwing her stone, by contrast, *is* an actual cause according to the restriction imposed by AC2(b). Given that Suzy throws her stone earlier than Billy, there is no way the actual causal process leading up to the bottle's shattering can be affected by changing Billy's actions.

et al., 2010; Rose, 2017). For our purposes the most relevant issue is that a successful descriptive account may still be explanatorily unsatisfactory. The definition given by Halpern and Hitchcock, for example, is already quite complex. Even if we suppose that the definition is descriptively adequate, there remains the question *why* we do and should employ such a complex causal concept.[7]

3 Functional Approaches and the Challenge of Purpose

The key idea of functional projects is to assume that causal reasoning can be useful for achieving various kinds of goals and then to understand and assess the concept with regard to these goals (Woodward, 2014, 2015, 2021).[8] Functional projects have been pursued in several ways. First, the functional approach has been employed as a heuristic for characterizing the actual use of the concept. In this case the functional project ultimately supports the descriptive project, for example, by suggesting relevant descriptive hypotheses, possible experiments for assessing these hypotheses, as well as interpretations of the experiments (Danks, 2013; Woodward, 2018). Second, there have been approaches that put a stronger emphasis on *evaluating* a concept based on its functions. According to Woodward, causal thinking can be understood as a "tool" or "technology" (2021, 30), and (like other tools and technologies) we can assess whether and to what degree it serves its purposes.

Here I will primarily be concerned with the evaluative approach. More specifically, I will explore the prospects of employing a functional approach to *improve* our concept of actual causation. Taking up Woodward's 'tool' metaphor I will investigate whether and under which circumstances the concept of actual causation can be engineered such that it serves its purposes better.

Like other proponents of functional accounts, I wish to delineate my approach from the metaphysical project of identifying what causation fundamentally is or what it reduces to. Instead, I am interested in the various ways concepts of actual causation can be made precise and to investigate which of those ways of making the concept precise are most useful.

The idea of improving our concepts is closely related to ideas of Carnapian explication and conceptual engineering. Carnap (1947) describes explicating a concept as the "task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept" (7f). Conceptual engineering is a more recent development, and it is closely related to Carnap's approach of explication. Cappelen (2018) identifies the following 'Master Argument' of conceptual engineering that underlies or motivates—at least implicitly—many revisionary projects in philosophy. First, words have certain meanings, but typically there are many similar

[7] Another potential problem for descriptive accounts is that in certain cases such as trumping cases (Schaffer, 2000), there seems to be disagreement among theorists even about the basic intuitions (Weslake, forthcoming). This kind of problem will not be relevant in the following because in our case of late preemption the initial intuition that Suzy is an actual cause but not Billy is uncontroversial.

[8] The focus here will be on functional approaches to causal concepts, but functional approaches also have been pursued with regard to other concepts, for example, 'knowledge' (Craig, 1990) and 'scientific explanation' (Woody, 2015).

meanings that the words could have instead. Second, we have no good reason to believe that the current meanings of our words are the best meanings that they could have. Third, we should make sure that our words have as good meanings as possible. This means that we (as philosophers) should assess and ameliorate those meanings.

But what does it mean to improve a concept? Reck (2012), for example, argues that Carnapian explications are not meant to be correct or incorrect. Instead, they can only be more or less fruitful. This immediately raises the question: "fruitful with respect to what goal?" (108) Likewise, when a conceptual engineer seeks to improve concepts, there is no unique standard to be followed. Instead, "improvement is relative to contextually specific purposes" (Cappelen, 2018, 137).

So in order to evaluate whether a suggested revision of the concept of actual causation is successful, we have to refer to a particular goal in a particular context. The concept of actual causation, however, is employed in a wide variety of contexts with many different goals in mind. It is employed, for example, when we aim to predict or explain some outcome, when we want to find out who is responsible for some outcome, and when we want to exert control over possible outcomes.

This gives rise to what I call the *challenge of purpose*. *Prima facie* one should not expect that those features that make the concept of actual causation particularly useful for one kind of goal will also be the features that make it useful for all other kinds of goals. This means that if we tune the concept of actual causation such that its performance with regard to some particular goal is improved, then that does not mean that it will also be improved with regard to all other potential goals. In fact, we will see that there are instances where attempts to improve the concept in one regard even have adverse consequences for the usefulness of the concept in other regards.

As indicated by the quotations above, it is widely accepted in the literature that a concept's functions or purposes play an important role for revisionary projects. It is also widely accepted that not appreciating the relevance of functions or purposes can lead to problems. Reck, for example, argues that a neglect of functions and purposes would be a "blind spot" in Carnap's methodology (2012, 97). The challenge of purpose is different from these existing discussions because it describes the more specific case where (1) the concept has several functions and (2) the functions imply conflicting criteria of success for potential revisions.[9]

In what follows I will focus on contexts that involve some form of late preemption, as described by the causal model in the foregoing section. I will show that in this kind of context tensions arise between two purposes of the concept of actual causation that are particularly salient: intervention and responsibility.[10] I will

[9] Are there other concepts where the challenge of purpose arises, besides the concept of actual causation? At least condition (1) seems to be fulfilled for a number of other concepts that are of central philosophical concern, such as knowledge ascriptions (Lossau, 2019) and truth. But it is unclear whether the different functions that have been ascribed to these concepts give rise to similar conflicts as in the case of actual causation. The challenge of purpose may also be particularly salient in revisionary projects that have a political dimension, such as Haslanger's (2000) revisionary account of the concepts of gender and race.

[10] Further tensions might arise with regard to other features of the concept. There has been discussion about whether and in what way causal judgment depends on considerations of normality and typicality,

proceed by first discussing the usefulness of the concept of actual causation from the perspective of intervention. I will take Halpern and Hitchcock's (2015) definition of actual causation as a template and assess each of the definition's conditions AC1, AC2, and AC3 from an interventionist viewpoint. We will see that each condition may under certain circumstances be motivated from this viewpoint except for condition AC2(b) (or other conditions that have been employed to delineate preempting and preempted factors in instances of late preemption). From the interventionist viewpoint there appear to be good reasons to depart from our initial causal intuitions and drop this condition, or at least alter it so that the definition better addresses the interventionist's goals. In Sect. 5 I shall turn to the viewpoint of responsibility. Again, I shall take Halpern and Hitchcock's definition as a template and assess each of the conditions AC1, AC2, and AC3. Here we will see that also condition AC2(b) is motivated in situations of late preemption, at least under some understanding of responsibility.

4 Intervention and Actual Causation

Most interventionist discussions so far have been primarily concerned with the structural causal relations exhibited by causal models. The fact that some variable X is causally related to another variable Y indicates that we may control Y by intervening on X. But what exactly is the relation between causation and manipulability? Woodward (2003) argues that there is no unique relation, and that instead we can define a range of causal concepts: total cause, direct cause, contributing cause, and actual cause. For example, if X is a total cause of Y, then there will be interventions on X that are sufficient to induce a change in Y. By contrast, if X is a contributing cause of Y, then it may be the case that a change in Y can only be achieved if an intervention on X is combined with further interventions.

Why think that claims of actual causation are at all relevant from the interventionist perspective? Shouldn't the causal dependencies summarized by a causal model be sufficient? Actual causation describes a relation between particular events or states of affairs that actually do take place (condition AC1). And these appear to matter for purposes of intervention, at least in some instances. We know that there are a wide range of potential causes of back pain, for example, little exercise, bad posture, or even a prolapsed disk or a broken spine. But if a particular patient visits her orthopedist, then the orthopedist will have to find out what actually causes *this patient's* back pain. This is relevant from the viewpoint of intervention. Knowing

Footnote 10 (Continued)

and how such a dependence would be explained (Hitchcock & Knobe, 2009; Alicke et al., 2011; Sytsma et al., 2012; Rose & Danks, 2012; Danks et al., 2014; Willemsen & Kirfel, 2019). In analogy to the present discussion one might ask whether and in what way the concept of actual causation *should* depend on such considerations (Blanchard & Schaffer, 2017; Fischer, 2021). Here the challenge of purpose might arise if this evaluation is sensitive to assumptions about the use of the concept in contexts of intervention or ascribing responisbility.

what actually causes the patient's back pain will in many cases determine the therapy that is to be prescribed.[11]

The relevance of condition AC3 can also be explained from an interventionist viewpoint. We want our interventions to target those factors that promise to realise our goals with minimal effort.

Our central concern here will be condition AC2. One might think that identifying Suzy as actual cause does not help with regard to saving the bottle. If we intervene on Suzy, the bottle will still be destroyed by Billy. So, it seems that what matters from the interventionist perspective is straightforward causal dependence rather than the kind of relationship described by condition AC2 and other approaches to actual causation (Hitchcock, 2013). But loosening the requirement of counterfactual dependence in the sense of condition AC2(a) still seems to be reasonable from the interventionist viewpoint. Even if there is no counterfactual dependence of $Y = y$ on $X = x$, then a relation of actual causation may still indicate that an intervention on $X = x$ can lead to $Y \neq y$ if we combine the intervention on X with further interventions on other factors $\vec{W}$. That is, the claim that Suzy is an actual cause is still helpful. It indicates that we may save the bottle if we combine an intervention on Suzy with interventions on other factors, in this case on Billy.

Moreover, the normality criterion imposes a useful restriction in many contexts. When an agent aims to prevent some outcome φ, she should be interested in strategies that address abnormal factors because such strategies are often easier to generalize (in the case of statistical norms) and tend to be in agreement with her normative commitments (in the case of moral norms)(Hitchcock & Knobe, 2009). The concrete strategies that an agent will come up with will depend on the context and her (at least partially) subjective normative commitments. What earlier seemed to be a potential drawback for the definition of actual causation now seems to be a benefit, at least from the pragmatic viewpoint of an agent who seeks to identify targets of intervention that suit her particular context and goals.[12]

Problems arise, however, when we try to justify condition AC2(b) from the viewpoint of intervention. In the stone-throwing case condition AC2(b) serves to distinguish preempting factors like Suzy (that are actual causes) from preempted factors like Billy (that are not actual causes). A causal explanation of the bottle's shattering that would not draw this distinction and treat Suzy and Billy on a par supposedly misses an important fact: namely that it was Suzy's and not Billy's stone that destroyed the bottle. But what does an intervening agent learn from this fact? Any agent that seeks to prevent the bottle's shattering needs to intervene on Suzy *and* Billy.

In order to see that extant functional approaches to the concept of actual causation struggle with this issue consider Hitchcock and Knobe's (2009) account. According

[11] Note, however, that the pain example involves enduring states. Condition AC1 will be more difficult to motivate from an interventionist perspective in examples involving irreversible causation, such as the preemption cases discussed here. For a discussion of the distinction see Ross and Woodward (forth).

[12] Note, however, that there are potential counterexamples. Suppose Suzy dies because a meteor hits her house. Suzy being home is normal, whereas a meteor strike is extremely abnormal. Yet the best strategy for saving Suzy would have been to make her leave her house, rather than intervening to stop the meteor. Thanks to an anonymous referee for suggesting this example.

to Hitchcock and Knobe, causes are difference-makers in a normalized version of the actual situation. In cases like the short-circuit/oxygen example discussed above this works. In the normalized version of the situation there, supposedly, is oxygen but no short circuit and, hence, no fire. In this normalized scenario the short circuit is a difference-maker while the oxygen isn't. According to Hitchcock and Knobe, this makes sense from the interventionist viewpoint: if we wish to prevent the fire, then it is much more convenient to avoid short circuits than to avoid oxygen.

But consider our example of late preemption. In the normalized version of this scenario the bottle, supposedly, remains intact because neither Suzy nor Billy throw their stones. In this normalized version Suzy's throwing her stone would be a difference-maker—which corresponds to the fact that Suzy is an actual cause. However, in the normalized situation Billy's throwing his stone is a difference-maker in just the same sense. So, while Hitchcock and Knobe's theory may help us understand the relevance of normality considerations from a functional perspective it does not explain why we distinguish between preempting and preempted factors.

Hitchcock (2017) also argues that claims of actual causation are useful because they track path-specific effects. We identify Suzy as actual cause because there is an active causal route linking her throwing the stone and the bottle's shattering. Hitchcock argues that knowledge of such an active route can be exploited if we want to save the bottle. In particular, he argues, it helps us identify the right combination of interventions needed to save the bottle: "In order to arrive at the right combination of interventions, […] I need to know that preventing Suzy from throwing will prevent the [bottle] from shattering, if I also intervene to prevent Billy from throwing" (124). I agree with Hitchcock that claims of actual causation thus can help identify potential targets for intervention.

It is not clear, however, how this explanation would help to justify a distinction between Billy and Suzy in our example of late preemption. In fact, if Billy had thrown his stone first and had preempted Suzy, then we would have to apply exactly the same combination of interventions. That is, in this regard knowledge of who, between Billy and Suzy, is the actual cause does not matter from the viewpoint of intervention and a concept that prioritizes Suzy seems to be questionable.

Another potential explanation for why we identify actual causes is that we have direct evidence of Suzy's causal role because we observe that Suzy's stone hits the bottle, and the bottle is shattered as a result. This immediately suggests that we should have intervened on Suzy in order to save the bottle. Billy's causal role, by contrast, is much less evident. Assessing Billy's causal role involves complex counterfactual considerations: we have to imagine what would have happened if Suzy had not hit the bottle. In particular, it may not be straightforward to evaluate whether Billy's throw would have had the sufficient momentum and would have been sufficiently accurate to destroy the bottle.

These epistemic considerations may serve as an explanation why we intuitively *do* prioritize the causal role of Suzy over the causal role of Billy. However, there is a different question of whether we *should* prioritize Suzy. The description of the example explicitly states that Billy's stone would have destroyed the bottle. This

establishes Billy's actions as a live threat to our goal of saving the bottle and implies that we should intervene on Billy just as we have to intervene on Suzy.[13]

So, from the interventionist perspective, what concrete adjustments should be pursued in response to these observations? Since this, of course, cannot be decided by looking only at a single case like late preemption, a conclusive answer is beyond the scope of this paper. But note the following. On the one hand, a conclusion that is suggested by the preemption case is to simply drop condition AC2(b), because it is responsible for delineating preempting and preempted factors. Interestingly, AC2(b) is also not needed in order to account for the majority of cases put forward in Halpern and Pearl's (2005) contribution, which was the first to propose AC2(b).[14] So, it seems like the late preemption scenario was the main motivation for this condition.

On the other hand, there are also cases that suggest that a restriction along the lines of AC2(b) is sometimes needed, also from the interventionist viewpoint. Suppose, for example, a prisoner D dies if A loads B's gun and B shoots or if C loads and shoots his gun. The structural equation is $D = 1$ iff $(A = 1 \wedge B = 1) \vee (C = 1)$ (Hopkins & Pearl, 2003). Suppose that in the actual case A does not load the gun, but C loads and shoots his gun. In this case we would not say that A is an actual cause of the victim's death. Our definition of actual causation reproduces this verdict only because $A = 0$ does not fulfil condition AC2(b). So, it seems, condition AC2(b) is doing important work in this case. Not identifying $A = 0$ as actual cause also seems to be plausible from the interventionist perspecitve. An agent who is interested in saving the prisoner should clearly target C but an intervention on A's actions is not needed to save the prisoner because B does not shoot anyway.[15] One possible conclusion that is indicated by these considerations is that one may need several different causal concepts in order to provide a satisfactory functional account from the interventionist viewpoint (Fischer (forthcoming), see also the outlook in Sect. 6 for further discussion).

So far, I have discussed Halpern and Hitchcock's (2015) definition of actual causation because their distinction between AC2(a) and AC2(b) makes the discussion particularly transparent, and my main concern has been with AC2(b). But what about definitions that do not involve AC2(b), such as Halpern's (2016) 'modified' definition? Like Halpern and Hitchcock's definition, this definition includes a condition that relaxes the requirement of counterfactual dependence. But Halpern's modified definition takes into account only scenarios in which variables $\vec{W}$ are kept fixed at their actual values (rather than set to non-actual values). Suzy is an actual cause,

[13] Note that one can even think of circumstances in which—for future purposes—focussing on Billy is more useful than focussing on Suzy. Consider a slight alteration of the original case in which Suzy preempted Billy in the observed scenario, but where, in general, Billy is the more reliable stone thrower. In this senario the judgement of actual causation would indicate that Suzy is a suitable target of intervention, but in order to save future bottles one should rather focus on Billy. Thanks to an anonymous referee for suggesting this example.

[14] More specifically, Halpern and Pearl's arsonist cases (both the conjunctive and disjunctive version), the forest fire case (Bennett, 1993), 'Billy's medical condition' (Hall, 2004), and double prevention (Hall, 2004) are all accounted for without AC2(b).

[15] Note, however, that some definitions including AC2(b) may encounter difficulties when considering a series of variations of this case as discussed by Beckers (2021).

according to this definition, because keeping fixed the fact that Billy does not hit the bottle ($BH = 0$), the bottle's shattering depends on Suzy's throwing the stone. By contrast, Billy is not an actual cause because there is no set of variables $\vec{W}$ such that if those variables are kept fixed in their actual values, there is a dependence of the bottle's shattering on Billy's throwing the stone.[16]

A potential objection to my argument here is that this establishes a significant difference between preempted and preempting factors. In order to prevent the bottle's shattering, we have to combine an intervention on Suzy with an intervention that keeps fixed the fact that Billy does not hit the bottle. An intervention on Billy, by contrast, would have to be combined with an intervention that changes other variables to non-actual values. This, one could object, is a significant difference because interventions that change variables to non-actual values represent a larger deviation from the actual setting and are more costly than interventions that simply hold fixed variables at their actual values.

But suppose we have intervened on Suzy such that she does not throw her stone or does not hit the bottle. Consider the concrete interventions that would then 'keep fixed' $BH = 0$: blocking the stone, taking away the bottle, ... Any such intervention would also be sufficient to change the value of BH to 0 if it had been 1 from the outset. This is related to the fact that the only reason for Billy's not hitting the bottle in the current scenario is that it has been destroyed by Suzy at the time when Billy's stone arrives at the initial location of the bottle. More specifically, interventions that 'keep fixed' $BH = 0$ would differ from interventions that change the value of BH from 1 to 0 only if they addressed the impact of Suzy's causal process on Billy's causal process. But this does not help because the only impact of Suzy's causal process on Billy's causal process is via the bottle's being destroyed.

In this section, I have highlighted an interesting tension between descriptive approaches to actual causation and the interventionist functional approach. On a descriptive level it is uncontested that we do distinguish preempting and preempted factors. Therefore descriptive accounts typically include a condition reflecting this distinction. From the perspective of the proposed functional approach, however, it seems like we should not endorse such a distinction—at least with regard to cases of late preemption. In particular, if causal concepts are supposed to indicate targets for intervention, then there is a clear sense in which preempted factors *should* be taken into account. A causal concept that tracks both preempting and preempted factors will be better suited to help us identify targets for intervention if our goal is to prevent outcomes such as the bottle's shattering.

[16] Other definitions of actual causation such as the definition by Hitchcock (2001) and the inferential account by Andreas and Günther (2021) allow a similar treatment of late preemption. Beckers and Vennekens (2018) propose a different treatment that takes into account the *timing* of the incoming stones. According to this account, Suzy produces the bottle's shattering and Billy does not because his stone arrives only after the bottle has been destroyed. But this does not make Suzy a better intervention target than Billy. In order to save the bottle an intervening agent would have to address both incoming stones independently of the details of their timing.

5 Responsibility and Actual Causation

In his pioneering treatment of actual causation in the framework of causal models Pearl identifies actual causation as the "ultimate criterion [⋯] for determining legal responsibility" (2000, 309). Since then, a large number of standard test cases have been drawn from legal contexts, and major theoretical insights have been inspired by the legal literature (such as Halpern and Hitchcock's notion of normality, which is inspired by Hart and Honoré (1959), among others). There have also been a number of attempts to employ the concept of actual causation to *define* concepts of responsibility and blame (Chockler & Halpern, 2004; Halpern, 2015, 2016).

In what follows I shall suggest a functional approach to the concept of causation from the perspective of responsibility. Assuming that the concept of actual causation has an important role to play in our practices of ascribing responsibility, what features should that concept of actual causation have in order to suit that purpose? As a template I will use again Halpern and Hitchcock's (2015) definition of actual causation.

Condition AC1 is clearly relevant for evaluations of responsibility. In legal inquiry it is typically not sufficient to speculate about all the many ways the occurrence of a kind of harm could be affected. Instead, we are interested in how a particular harm did in fact come about. This can only be decided on the basis of claims about which events did in fact occur in the situation under consideration. On the level of causal models, therefore, the relevant kind of causal relations are relations between variables that take on particular values, as in '$X = x$ is an actual cause of $Y = y$.'

AC3 was introduced by Halpern and Pearl (2005) as a minimality condition. For example, they argue that if striking a match qualifies as an actual cause according to conditions AC1 and AC2, then under certain circumstances striking a match plus sneezing would also qualify as actual cause. The minimality condition excludes such irrelevant details. A concept of actual causation employed for determining an agent's responsibility should clearly fulfil this criterion in order to help us indicate as specifically as possible what an agent is to be held responsible for.

Next, we shall turn to condition AC2. In legal contexts questions of causation are often addressed with the 'but-for' test, which requires straightforward counterfactual dependence of the outcome on the actual cause. Suzy is not a but-for cause of the bottle's shattering because Billy would have destroyed the bottle if Suzy had not. But note how little of an excuse it would be if Suzy defended herself claiming that if she had not destroyed the bottle, then Billy would have. The morally appropriate reaction to Suzy's actions is to hold her responsible. The concept of but-for cause would not serve the purpose of indicating this. A concept of causation involving condition AC2(a), by contrast, would fulfil the purpose of indicating this and, thus, seems to be better suited.

Moreover, employing a restriction along the lines of the normality criterion seems to be well motivated in contexts concerned with responsibility. It should

be noted that in the literature on causation in the law there has been major disagreement about the exact role that normality considerations play in legal inquiry and, in particular, at which stage of legal inquiry normality-related considerations are legitimately introduced. Hart and Honoré, for example, argue that a normality criterion like the one proposed by Halpern and Hitchcock is an integral part of the legal concept of causation. On the other hand, theorists standing in the tradition of legal realism, such as Richard Wright (1985; 2001), argue that the legal concept of actual causation should not be seen as depending on considerations of normality. But theorists like Wright also agree that normality considerations are relevant in assessing an agent's responsibility (even if they do not play a legitimate role in the part of legal inquiry that is concerned with questions of actual causation).

Let us now turn to condition AC2(b). There is a natural understanding of the situation on which Suzy is responsible for the bottle's shattering while Billy isn't. After all it was Suzy's stone that hit and destroyed the bottle. But does this explain the relevance of condition AC2(b)? A potential objection here is that Billy is just as blameworthy as Suzy because he attempted to destroy the bottle. In the following we will see that the relevance of actual causation (and in particular the relevance of condition AC2(b)) depends upon one's understanding of responsibility.

Suppose Suzy and Billy are assassins that both attack an innocent victim. They have been sent on their missions by different and independent clients and they attack without knowing of each other. Suzy and Billy pull the triggers of their guns almost at the same time but Suzy shoots a little earlier and her bullet instantly kills the victim. Does it matter that Suzy (but not Billy) is the actual cause in this kind of case?

This depends on one's theory of responsibility.[17] I will not provide a systematic review of such theories but let me sketch two kinds of theories and possible ways in which these theories could account for the example. First, consider moral influence theories. According to Schlick (1939), for example, the aim of imputing responsibility to a person is punishment (or reward). Punishment, in turn, "is an educative measure, and as such is a means to the formation of motives, which are in part to prevent the wrongdoer from repeating the act (reformation) and in part to prevent others from committing a similar act (intimidation)" (152).[18]

The key concern of moral influence accounts is to prevent future harm. Imposing blame is seen as a method of corrective *intervention* that discourages the agent from repeating her undesired action and deters others from similar actions. In this sense moral influence theories are closely analogous to interventionist theories

[17] The following discussion is closely related to considerations about (resultant) moral luck (Williams (1981); Nagel (1979)). Proponents of moral luck argue that blameworthiness depends on factors outside the agent's control. Suzy and Billy both perform the same action (pulling the trigger of the gun) with the same intention (killing the victim). As luck would have it, however, Suzy shoots a split second earlier than Billy. For adherents of moral luck, the fact that Suzy's bullet did in fact kill the victim makes her more blameworthy than Billy. Drawing a distinction between preempted and preempting causes would seem to undergird this intuition. By contrast, opponents of moral luck argue that factors beyond the agent's control do not (or should not) have such an influence on our moral evaluation.

[18] Further accounts along these lines have been provided by J. J. C. Smart (1961), Daniel Dennett (1984), and Manuel Vargas (2008).

of causation, even if corrective intervention is not to be understood in the technical sense of intervention proposed, for example, by Woodward (2003). But if our goal is to prevent future assassinations, then we certainly should discourage both Suzy's and Billy's attempts of assassination—independently of whether they were preempted or not. This means that from the perspective of moral influence there appear to arise similar difficulties for motivating a condition like AC2(b) as those difficulties discussed in the foregoing section. This result should not be surprising if one takes the analogy between moral influence theories and an interventionist understanding of causation seriously.

Retributivists, by contrast, take the actual wrong done by an agent to be an independent desert basis for punishment. Proponents of retributive justice hold (1) that those who commit a wrongful act morally deserve to be punished proportionately, (2) that it is intrinsically morally good if wrongdoers receive the punishment they deserve, and (3) that it is not permissible intentionally to punish the innocent (Walen, 2021). For example, Moore argues that "two persons, each of whom executes the same voluntary act with the intention of killing an innocent victim and do so with equal lack of excuse, differ in their deserved punishments if they differ in the amount of wrong they each succeed in doing" (1994, 237). More specifically, Moore argues that the person who succeeds in killing the innocent victim deserves a punishment that is more severe than the one deserved by the person that fails.

Retributivists often rely on direct intuitive support when they justify punishment. Moore, for example, relates the larger blameworthiness of the successful agent to a first-person experience of guilt. He argues that "we experience greater guilt when we have caused some harm that we [...] tried to cause [...] than we experience when we have been equally culpable but we have not caused such a harm" (2009, 30). In support of this Moore refers to the relief we typically feel in cases of near misses: "we know we just escaped something, namely being culpable killers of an innocent." (ibid.) There is an extensive discussion on whether experience of guilt lends any support to the claim that agents are more blameworthy when they cause some harm. Moreover, there is a range of attempts to provide alternative explanations for the apparent feeling of guilt in response to caused harm (see, for example, Wolf (2000); Enoch and Marmor (2007); Domsky (2004) and the responses in Moore (1994) and (2009, 30–3)). I will not try to settle this debate here. For our purposes it suffices to say that Moore's account represents an understanding of responsibility that appears to be deeply engrained in our practices of ascribing blame.

Moore argues that a similar kind of reasoning applies to situations involving late preemption. More specifically, Moore suggests that if one is the preempted factor in such a situation (like Billy in our example) that one's appropriate reaction is "still one of relief at a near miss: 'I almost did a great wrong, but as luck would have it, I didn't—someone (or something) else did'" (2009, 30). That is, even though Billy attempted to kill the innocent victim and the victim did in fact die, feeling relief is still an appropriate reaction for Billy. The reason, according to Moore, is an underlying principle of ownership, according to which Suzy (but not Billy) 'owns' the victim's death. So, according to Moore the morally appropriate attitudes towards preempters are different than those towards the preempted. If the function of the

concept of actual causation is to track this difference, then that concept should include a condition reflecting this difference.

A potential worry here is that my approach represents an attempt to provide a reductive account of causation in terms of responsibility and that any such attempt is bound to fail. In order to make progress, one would need to show that the underlying concept of responsibility is less troublesome than the concept of causation and that the concept of responsibility does not rely on a presupposed understanding of causation.[19]

I take it to be an open question whether such a reductive account is feasible. Moral influence theorists justify ascriptions of responsibility with a view to preventing future harms. The associated concept of prevention is a causal concept. So, a reductive account along these lines may face difficulties. Retributivists justify ascriptions of responsibility with a view to experiences of feelings of guilt or relief. Moore appears to relate such experiences to prior causal judgments. Alternativley, such experiences could be understood as basic. But whether they provide a sufficient basis for a reductive account of responsibility is questionable.

My functional approach does not attempt to provide a reductive account. Instead, my strategy has been to explore potential constraints on a concept of actual causation that would make it a useful tool in ascriptions of responsibility. My starting point has been to suppose that it is an important purpose of our causal concepts to help us indicate responsibility. What are the conditions that a concept of causation needs to fulfil to suit that purpose? I have argued that this depends on one's understanding of responsibility. Without taking sides, I have shown that on a moral influence theory of responsibility one may face difficulties motivating a condition like AC2(b), difficulties that are similar to those faced by interventionists. On a retributivist theory, however, a condition like AC2(b) appears to do important work.

Let me sum up. One's take on the difference between preempting and preempted actions in morally laden situations will depend upon one's take on the concept of responsibility. On a retributivist account the difference is important, meaning that a concept of actual causation that includes condition AC2(b) fulfils an important function, especially in some contexts involving late preemption. On this account—and in contrast to the interventionist perspective developed in the foregoing section—our intuitions associated with actual causation should *not* be revised in this regard.

6 Conclusion and Outlook

In this paper I have argued that functional approaches to actual causation—at least when they are pursued with the goal to improve our concepts—face a challenge of purpose. The idea of such functional approaches is to ask: what kind of causal concept would be most conducive to our goals? A central result of my discussion is that the answer to this question may depend on the context of application and

[19] This is analogous to an objection faced by interventionist accounts: Woodward (2003) spells out concepts of causation employing a concept of intervention. But the concept of intervention presupposes some preliminary understanding of causation.

our assumptions about those goals. In my discussion I have focused on a scenario involving late preemption. I have argued that if our primary goal is to facilitate intervention, then a causal concept that does not involve condition AC2(b) or other conditions distinguishing between preempting and preempted factors may be the most useful concept in this context. If our primary goal, however, is to facilitate our practices of ascribing responsibility, then the assessment of condition AC2(b) has to be pursued much more carefully because, depending on one's understanding of responsibility, the difference between preempting and preempted factors may indeed be relevant.

What are the consequences of the challenge of purpose for functional approaches to actual causation? One consequence is that revisionary accounts should not focus on just one of the multiple functions that causal concepts supposedly fulfil because this may lead to a biased outcome. In this light the current dominance of interventionist approaches may appear problematic. At the same time, I do not think that the challenge poses an insurmountable problem to functional projects. But the challenge of purpose may indicate that functional approaches are unlikely to provide a unified concept of actual causation. Instead, I take the challenge of purpose to provide tentative support for pluralist theories of actual causation.[20]

This is an interesting result because even though there is a variety of pluralist theories of (actual) causation (Hitchcock, 2003; Hall, 2004; Cartwright, 2007), such pluralist theories are largely motivated by considerations associated with the descriptive project. Surveying a large sample of example cases, Hall (2004), for example, argues that counterfactual accounts fail to provide a univocal concept of causation because they are at odds with the transitivity, locality, and intrinsicness of causation. He concludes that besides a dependence notion of causation (which can be spelled out in terms of counterfactual conditionals) we need a causal notion of production.

My discussion of the challenge of purpose provides additional support for pluralist theories of actual causation from a functional perspective. Suppose we were able to provide a unified concept of actual causation that is perfectly adequate from a purely descriptive point of view. Then it may still be the case that such a concept's performance can be improved. I have shown that what counts as an improvement sometimes depends on the particular purpose. Different purposes may require alterations of the concept that pull in opposing directions. This means that in order to facilitate all purposes optimally we would have to give up the initial unified concept and endorse a plurality of causal concepts instead.

Acknowledgements Many thanks for helpful discussion and comments to Hasok Chang, Mathias Frisch, Céline Henne, Christopher Hitchcock, Ruward Mulder, Miguel Ohnesorge, Tammo Lossau, Bobby Vos, Oscar Westerblad, and Lucie White.

[20] There are different kinds of pluralist theories of causation. According to a distinction proposed by Hitchcock (2007b), extramural pluralists argue that we need to employ more than one kind of basic building block (laws, counterfactuals, physical processes...) to make sense of causal concepts, as in the theories of Hall (2004) and Cartwright (2007). Intramural pluralists argue that one kind of basic building block may serve to define more than one concept of causation, as proposed by Hitchcock (2003). My argument lends support to the idea that there should be more than one concept of actual causation. Whether more than one kind of basic building block is needed or not remains open.

Author Contributions Not applicable.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) (Grant Number 254954344/GRK2073)

Availability of Data and Material Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest None.

References

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *The Journal of Philosophy, 108*(12), 670–696.

Andreas, H., & Günther, M. (2021). A Ramsey test analysis of causation for causal models. *The British Journal for the Philosophy of Science, 72*(2), 587–615.

Beckers, S. (2021). Causal sufficiency and actual causation. *Journal of Philosophical Logic, 50,* 1341–1374.

Beckers, S., & Vennekens, J. (2018). A principled approach to defining actual causation. *Synthese, 195,* 835–862.

Bennett, J. (1993). Event causation: The counterfactual analysis. In Sosa, Ernest and Tooley, Michael, editors, *Causation*, pages 217–33. Oxford University Press, Oxford.

Blanchard, T., & Schaffer, J. (2017). Cause without default. In H. Beebee, H. Price, & C. Hitchcock (Eds.), *Making a Difference* (pp. 175–214). Oxford University Press.

Cappelen, H. (2018). Conceptual engineering: The master argument. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual ethics and conceptual engineering* (pp. 132–51). Oxford University Press.

Carnap, R. (1947). *Meaning and necessity*. Chicago University Press.

Cartwright, N. (2007). *Hunting causes and using them*. Cambridge University Press.

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of A.I. Research, 20,* 93–115.

Craig, E. (1990). *Knowledge and the state of nature. An essay in conceptual synthesis*. Clarendon Press.

Danks, D. (2013). Functions and cognitive bases for the concept of actual causation. *Erkenntnis, 78*(1), 111–28.

Danks, D., Rose, D., & Machery, E. (2014). Demoralizing causation. *Philosophical Studies, 171,* 251–277.

Dennett, D. C. (1984). *Elbow Room*. MIT Press.

Domsky, D. (2004). There is no door: Finally solving the problem of moral luck. *Journal of Philosophy, 101,* 445–64.

Enoch, D., & Marmor, A. (2007). The case against moral luck. *Law and Philosophy, 26,* 405–36.

Fischer, E. (2021). Causation and the problem of disagreement. *Philosophy of Science, 88*(5), 773–783.

Fischer, E. Three concepts of actual causation. *British Journal for the Philsophy of Science*, forthcoming.

Gallow, J. D. (2021). A model-invariant theory of causation. *The Philosophical Review, 130*(1), 45–96.
Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., et al. (2010). Actual causation: A stone soup essay. *Synthese, 175*(2), 169–192.
Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and counterfactuals* (pp. 225–76). MIT Press.
Halpern, J. Y. (2015). Cause, responsibility and blame: a structural-model approach. *Law, Probability and Risk, 14*(2), 91–118.
Halpern, J. Y. (2016). *Actual Causality*. Cambridge, MA: MIT Press.
Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science, 66,* 413–57.
Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science, 56*(4), 843–87.
Hart, H. L. A., & Honoré, A. M. (1959). *Causation in the Law*. Clarendon Press.
Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs, 34*(1), 31–55.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy, 98*(6), 273–99.
Hitchcock, C. (2003). Of Humean Bondage. *The British Journal for the Philosophy of Science, 54,* 1–25.
Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review, 66,* 495–532.
Hitchcock, C. (2007). How to be a causal pluralist. In P. K. Machamer (Ed.), *Thinking about causes: From Greek philosophy to modern physics* (pp. 200–221). University of Pittsburgh Press.
Hitchcock, C. (2013). What is the "cause'' in Causal Decision Theory. *Erkenntnis, 78,* 129–146.
Hitchcock, C. (2017). Actual causation: What's the use. In H. Beebee, C. Hitchcock, & H. Price (Eds.), *Making a difference* (pp. 116–31). Oxford University Press.
Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy, 106*(11), 587–612.
Hopkins, M. and Pearl, J. (2003). Clarifying the usage of structural models for commonsense causal reasoning. *Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, pages 83–9.
Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*(17), 556–67.
Lossau, T. (2019). Three things to do with knowledge ascriptions. *Episteme*, pages 1–12.
Moore, M. (1994). The independent moral significance of wrongdoing. *Journal of Contemporary Legal Issues, 5,* 237–81.
Moore, M. (2009). *Causation and responsibility: An essay in law, morals, and metaphysics*. Oxford University Press.
Nagel, T. (1979). *Mortal Questions*. Cambridge University Press.
Pearl, J. (2000). *Causality. Models, reasoning, and inference*. Cambridge University Press.
Reck, E. (2012). Carnapian explication: A case study and critique. In P. Wagner (Ed.), *Carnap's ideal of explication and naturalism* (pp. 96–116). Palgrave Macmillan.
Rose, D. (2017). Folk intuitions of actual causation: A two-pronged debunking explanation. *Philosophical Studies, 174,* 1323–1361.
Rose, D., & Danks, D. (2012). Causation: Empirical trends and future directions. *Philosophy Compass, 7*(9), 643–653.
Ross, L. N. and Woodward, J. Irreversible (one-hit) and reversible (sustaining) causation. *Philosophy of Science*, forth.
Schaffer, J. (2000). Trumping preemption. *The Journal of Philosophy, 97*(4), 165–181.
Schlick, M. *Problems of Ethics*. Prentice-Hall, Inc., New York, 1939. Authorized translation by David Rynin.
Smart, J. J. C. (1961). Free-will, praise and blame. *Mind, 70*(279), 291–306.
Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. Springer-Verlag.
Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: Rethinking th role of statistical typicality in ordinary causal attributions. *Studies in history and philosophy of biological and biomedical sciences, 43,* 814–820.
Vargas, M. (2008). Moral influence, moral responsibility. In N. Trakakis & D. Cohen (Eds.), *Free will and moral responsibility* (pp. 90–122). Cambridge Scholars Press.
Walen, A. (2021). Retributive Justice. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Weslake, B. A partial theory of actual causation. *The British Journal for the Philosophy of Science*, forthcoming.
Willemsen, P., & Kirfel, L. (2019). Recent empirical work on the relationship between causal judgements and norms. *Philosophy Compass, 14*(1), e12562.
Williams, B. (1981). *Moral Luck*. Cambridge University Press.
Wolf, S. (2000). The moral of moral luck. *Philosophical Exchange, 31,* 5–19.
Woodward, J. (2003). *Making things happen*. Oxford University Press.
Woodward, J. (2014). A functional account of causation; or, a defense of the legitimacy of causal thinking by reference to the only standard that matters-usefulness (as opposed to metaphysics or agreement with intuitive judgment). *Philosophy of Science, 81*(5), 691–713.
Woodward, J. (2015). Methodology, ontology, and interventionism. *Synthese, 192,* 3577–99.
Woodward, J. (2018). Normative theory and descriptive psychology in understanding causal reasoning: The role of interventions and invariance. Epistemic Studies. In W. J. Gonzalez (Ed.), *Philosophy of Psychology: Causality and Psychological Subject* (pp. 71–104). De Gruyter.
Woodward, J. (2021). *Causation with a human face. Normative theory and descriptive psychology*. Oxford University Press.
Woody, A. I. (2015). Re-orienting discussions of scientific explanation: A functional perspective. *Studies in History and Philosophy of Science, 52,* 79–87.
Wright, R. (1985). Causation in tort law. *California Law Review, 73*(6), 1735–828.
Wright, R. (2001). Once more into the bramble bush: Duty, causal contribution, and the extent of legal responsibility. *Vanderbilt Law Review, 54*(3), 1–51.