

Singular Causation

David Danks

Abstract

Causal relations between specific events are often critically important for learning, understanding, and reasoning about the world. This chapter examines both philosophical accounts of the nature of singular causation, and psychological theories of people's judgments and reasoning about singular causation. It explores the content of different classes of theories, many of which are based on either some type of physical process connecting cause and effect, or else some kind of difference-making (or counterfactual) impact of the cause on the effect. In addition, this chapter examines various theoretical similarities and differences, particularly between philosophical and psychological theories that appear superficially similar. One consistent theme that emerges in almost every account is the role of general causal relations in shaping human judgments and understandings about singular causation.

Key Words: singular causation, physical process, difference-making, reasoning, counterfactual, effect

Singular Versus General Causation

In many people, caffeine causes slight muscle tremors, particularly in their hands. In general, the *Caffeine* → *Muscle Tremors* causal connection is a noisy one: someone can drink coffee and experience no hand tremors, and there are many other factors that can lead to muscle tremors. Now suppose that Jane drinks several cups of coffee and then notices that her hands are trembling. An obvious question is, did *this* instance of coffee drinking cause *this* instance of hand trembling? Structurally similar questions arise throughout everyday life: Did *this* pressing of the “k” key cause *this* change in the pixels on the computer monitor? Did *these* episodes of smoking cause *this* lung cancer? Did *this* studying cause *this* test score? And so on. These questions all ask about singular causation in a particular situation, in contrast with general causation across multiple cases. They are thus particularly salient in situations in which we care about that specific case, as in many legal contexts,

social interactions, physical explanations of anomalous events, and more.

Singular causation is cognitively challenging, as it is quite unclear how we could come to know that, say, this coffee drinking caused these muscle tremors. On the one hand, we cannot directly observe singular causation, as famously noted by David Hume (1748/2001), but instead observe only sequences of events. One might hope that so-called causal perception (see also White, Chapter 14 in this volume) could provide a way to directly observe singular causation, but there are reasons to suspect that causal perception is not just the straightforward observation of singular causation. In particular, the particular percepts must instead be understood or interpreted in light of background causal knowledge. On the other hand, we cannot use statistical or other inductive inference methods to learn directly about which singular causation relations obtain because we have (by definition) only a single case. There are many different methods for inferring

causal structure from observations, but those all require multiple cases, so are not directly applicable for singular causation. Moreover, even if they were usable, the causal relations across multiple situations need not perfectly track the causal relations in any particular case. For example, it is perfectly consistent for caffeine to cause shaky hands, but for *these* particular muscle tremors to be caused by *this* dose of medication, rather than the coffee that one just consumed.

Judgments about singular causation require more than just knowledge of the general causal relations that obtain in some domain; at the least, we also need to know what events actually occurred in this specific case. That is, singular causation requires knowledge of (at least) both the general causation that applies in cases of this type, as well as details about this particular case (Hitchcock, 2012). In fact, as we see in the following section, “Singular Causation, Normatively,” knowledge of singular causation may well require even more information (e.g., about particular physical relations, or default values). The key observation here is simply that we cannot divorce discussions of singular causation from general causation: the latter is critical for the former, even if only because those are the relations that can possibly combine in a singular fashion.

Determining singular causal relations is a kind of causal reasoning, though one with a very specific goal. More generally, we must recognize that pragmatic or contextual factors can often make a difference in how we think about singular causal relations. For example, if a lit match is dropped into a basket of paper, then it seems quite clear that the match is a singular cause of the paper burning. It is less clear how to think about the oxygen in the room; it is certainly causally relevant, but it also seems less central than the match, at least in a pragmatic sense. Those intuitions might well change in a different context: the presence of oxygen is more plausibly a singular cause of the fire if we are on the International Space Station. The normative and descriptive theories of singular causation discussed in this chapter all address this distinction, though they do so in quite different ways. The key here is simply that we do draw a distinction between causes and enabling or background factors, whether that distinction is based on our pragmatic interests or more objective features of the situation or context.

There is a long history of philosophical and normative inquiry into the nature of singular causation, and so this chapter begins there (the section “Singular Causation, Normatively”). The theories

that have been proposed can be roughly divided into two different types, though both face significant questions and concerns. Those two types of philosophical theories appear to correspond nicely to two types of descriptive, psychological research on singular causation, though as we see in section “Singular Causation, Descriptively,” neither of the normative, philosophical theories has clear, unequivocal empirical support. Moreover, there has been relatively less experimental study of singular causation judgments, and so there are many aspects of singular causation judgments that we do not yet understand. The normative, philosophical accounts thus suggest avenues for potentially fruitful future experiments. I conclude the chapter, in the section “Open Problems and Challenges,” with discussion of some key open problems and challenges. Before beginning, a terminological note: the contrast between causation in a specific instance versus across multiple instances is also sometimes described as “token versus type causation” (the particular cases involve “tokens” of the relevant “types” across situations) or “actual versus potential causation” (in a particular situation, only some of the “potential” causal relations “actually” obtain). I will use the language of “singular versus general causation” in this chapter, but readers should be aware that these ideas and theories are also discussed under other names.

Singular Causation, Normatively

Normative theories of singular causation aim to provide an account of which factors were actually the causes of some event, rather than which factors are typically picked out by experimental participants. This focus presents a verification challenge, however: How can we test our theories of singular causation, if not against human responses? The basic strategy in this area is to compare the judgments of a proposed theory against solid, stable intuitions in critical test cases. We test our normative theories against our considered, reflective judgments about key situations, and reject those theories that fail to match on those key points. For example, if I smash a glass with a hammer (and there are no unusual factors at play), then any theory of singular causation should conclude that my hammer strike caused the breaking of the glass. In addition, we might argue that our theories should exhibit particular higher-level features; for example, we might require that only distinct, separate events can singular cause one another (Menzies, 1996). In practice, however, testing against key examples is the dominant way of

arguing for a particular normative theory of singular causation. Importantly, not just any intuitions count, but only those that are widely held, based on careful thought, and so forth. This strategy is analogous to one natural way (though not the only way) to develop a normative theory of formal logic. One would not want to build a theory of formal logic by considering everyday judgments about whether some argument is logically valid, particularly given what we know about the factors that influence people's judgments on these matters. Instead, our theory must conform to those inferences that we agree, upon careful reflection and consideration, are truth-preserving (e.g., *modus ponens*: the combination of the two premises "If *A*, then *B*" and "*A*" necessarily implies "*B*").

There are two dominant types of theories of singular causation—one based on physical processes that connect the cause with the effect (discussed in the subsection "Physical Process Approaches") and one based on the cause making a difference to the effect (the subsection "Difference-Making Approaches"). Roughly, the first type says that *C* is a singular cause of *E* just when there is an appropriate physical process connecting *C* and *E*, while the second type holds that singular causation depends on whether *E* would have been different if *C* had been different (in the right way, in the right possible world). Both types of theories naturally explain some, but not all, stable judgments, and much of the normative, philosophical research on singular causation has aimed to find better formalizations of the underlying intuitive ideas so that the theories can capture more judgments. Some of these elaborations are described in the next two subsections, but the details are arguably less important than the overall intuitions: either singular causes are those factors that physically produce the effect, or they are the events that made an actual difference in the effect.

It will be helpful to have a running example for the remainder of this chapter; I borrow one from Hall (2004), though many others would do. Suppose we have two children, Suzie and Billy, who are throwing rocks at bottles in an abandoned lot. Both Suzie and Billy are excellent throwers who are, for all practical purposes, perfectly accurate: if one of them throws a rock at a bottle, then the rock hits and shatters that bottle. That is, the general causal relations are "Suzie's throws cause bottles to break" and "Billy's throws cause bottles to break." Now suppose that, on one particular occasion, both Suzie and Billy throw rocks at the very same bottle, but

that Suzie throws slightly earlier so that her rock is the one that hits the bottle (and the bottle subsequently breaks). A strong, obvious intuition is that Suzie's throw is a singular cause of the bottle breaking, and Billy's throw is not. Any normative theory of singular causation should presumably capture this clear intuition, and we now see two different ways of doing so.

Physical Process Approaches

One natural view about singular causes is that they are just the events that actually physically influence the outcome. The lit match is a singular cause of the fire because the match actually physically transmits the necessary energy to the paper. Suzie's throw was a singular cause of the bottle breaking because it imparted more force than could be absorbed by the bottle structure. This intuition finds expression in physical process (normative) theories of singular causation. The general idea is that *C* is a singular cause of *E* just when *C* changes *E* through a physical/causal process connecting the two (see also Johnson & Ahn, Chapter 8 in this volume). Of course, we must provide some independent characterization of a "physical/causal process," or the account of singular causation will be viciously circular. One of the first precise theories was Wesley Salmon's (1984) mark transmission theory. In this theory, the world is composed of processes that exhibit reasonably stable structure over some period of time, and so "marks" (i.e., structural modifications) made on those processes can persist over time. A particular causal interaction occurs when a mark from one causal process is transmitted to a different process. Singular causation more generally, then, consists of propagation and mark transmission in these processes. For example, Suzie's rock caused the bottle to break because the causal process corresponding to her thrown rock interacted with the causal process corresponding to the intact bottle to yield a "mark" on the bottle, as the bottle's structure was modified (to the "broken" state). In contrast, Billy's rock had no such interaction with the causal process of the bottle, and so it was not a singular cause of the bottle breaking, even though it would have broken the bottle if Suzie's rock had not hit first.

The mark transmission theory can explain many standard intuitions, but suffers from significant ambiguities in key notions (Kitcher, 1989). For example, processes are supposed to be those spatio-temporal regions ("world lines") that exhibit a degree of structural uniformity or consistency when not

interacting with other processes. This restriction is necessary to ensure that processes are appropriately coherent; random spatiotemporal regions ought not be considered causal processes. The problem is that this requirement is arguably too strong: for example, there are many seeming causal processes that require constant interaction with a background of other processes (e.g., sound waves moving through a medium), and so fail to meet the condition that they continue in the absence of interactions with other processes. Mark transmission theory thus rules out factors that are clearly (intuitively) singular causes. At the same time, mark transmission theory is overly permissive, as standard ways of defining a “mark” imply that many processes that are *not* intuitively causal can nonetheless transmit a “mark.” In particular, a “mark” cannot simply be a change in a property, since there are many property changes that are not true structure modifications. We need some way of precisely saying which changes in which properties count as “marks,” and no satisfactory account has been offered.

These problems with the mark transmission theory suggest that we should instead ground our physical process theory of singular causation even more in our best theories of physics. Conserved quantity theories (Dowe, 1992, 2007; Salmon, 1997) do exactly that. At a high level, the intuition is that causal processes are those that have a conserved quantity (e.g., mass-energy, momentum) and causal interactions are exchanges of that conserved quantity. Suzie’s rock, for example, has a certain mass-energy while in flight,¹ and then transfers some of that mass-energy to the bottle, which leads to it breaking; Billy’s rock is also a causal process with conserved mass-energy, but it does not interact similarly with the bottle, so is not a singular cause of the bottle breaking. Crucially, the conserved quantity theories all hold that which quantities are *actually* conserved is something that our best scientific theories aim to discover. The (singular) causal relations in the world are an objective matter; our best sciences help us learn which quantities actually are the conserved ones. And of course, it may be a quite difficult task to actually discover the physical processes that underlie some complex causal relation, particularly in sciences other than physics. For the physical process proponent, though, it is important to keep this epistemological challenge separate from the metaphysical question of what (singular) causation is.

The obvious concern about physical process theories is that we seem to sometimes have a

singular causal relation without any relevant process or exchange of a conserved quantity connecting the putative cause and effect. For example, if I fail to water my plants, then it certainly seems that the lack of water causes their death, even though there is no (relevant) causal process between me and the plants. In fact, we think that the absence of such a process is exactly what causes the problem! More generally, many canonical cases of singular causation seem to involve the absence or removal of a causal factor; in these cases, there is no physical process between cause and effect, and thus no actual causation according to physical process theories (Schaffer, 2000). The standard reply is to argue that these cases correspond to “quasi-causation,” a relation that looks and behaves much like singular causation, but is not actually causation. More precisely, Dowe (2001) argues that quasi-causation is grounded in the truth of key counterfactuals about what would have happened if, for example, the missing causal factor had actually been present. For example, lack of water (quasi-)causes my plants’ death because of the truth of the counterfactual “if I had watered them, then the watering would have caused them to live,” where “caused” in this counterfactual is understood in the standard physical process sense. Quasi-causation relations are thus “causal” relations, but depend on counterfactual causation rather than actual causation.

Appeals to quasi-causation can help to explain many of our strong intuitions about singular causal relations in the world. At the same time, the use of quasi-causation comes at a cost for physical process theories. Part of the intuitive appeal of those theories is that they enable us to understand singular causation entirely in terms of the actual world. The only things that matter for these theories are the processes and interactions that actually occur (and are thereby observable, testable, and so on). In contrast with the difference-making theories discussed in the next subsection, we do not need to consider what would have happened if the world had been different in certain ways. Quasi-causation does not have this appealing feature, however, as it depends critically on one or more counterfactuals. Thus, the physical process theorist is in the difficult position of either (a) concluding that cases of prevention or causation by omission are not on par with other causal relations; or (b) embracing counterfactuals and so losing an appealing feature of the theories (see also McGrath, 2005). Option (a) appears to rest largely on an intuition that omissions and preventions are not “real” causes, but there is little

theoretical justification or experimental support for that intuition. Option (b) faces the dual challenges of explaining both which counterfactuals are relevant, and also their truth-conditions. Physical process theories are arguably ill-equipped to handle either challenge, in large part precisely because they are grounded in the actual world (Schaffer, 2001). If we are going to use counterfactuals in our analysis of singular causation, then we should perhaps instead start with ones about how the effect would be different if various factors had varied.—that is, we should consider a difference-making approach.

Difference-Making Approaches

The idea that the singular causes are those that made a difference to the effect dates back at least to Hume (1748/2001), though he is better known for his associationist understanding of causation. The fundamental challenge with grounding singular causation in difference-making is that any such analysis must necessarily be counterfactual in nature. A factor makes a difference only if the world would have been different (in the right way) if the factor had been different, but we do not have direct access to the relevant counterfactual scenarios. For example, the claim “this hammer strike made a difference in this glass shattering” implies that the glass would not have shattered if the hammer had not struck it, but of course the hammer did strike it. Thus, any difference-making theory of singular causation must depend on counterfactuals, which can be difficult to assess. Moreover, the difference-making theory must also say *which* counterfactuals are relevant. In the hammer/glass example, the relevant counterfactual is obvious, but other scenarios are much less clear. Consider again the case of the perfectly accurate Billy and Suzie throwing rocks at a bottle, where Suzie’s rock is the one that actually makes contact. In this case, the “obvious” counterfactual—if Suzie had not thrown her rock, then the bottle would not have broken—turns out to be false (since Billy’s rock would have broken the bottle instead), even though (by assumption) Suzie’s throw is the singular cause of the bottle breaking. The relevant counterfactual needs to involve Billy, as Suzie’s difference-making can only be seen when he refrains from throwing his rock. The key challenge for a difference-making theory of singular causation is thus to explain which counterfactuals ground the singular causal claim.

David Lewis (1973, 2000) based his answer to this question on the notion of possible worlds. More specifically, the relevant counterfactual is

determined by the closest possible world in which the potential singular cause *C* did not occur. *C* is an actual singular cause just when, in this closest possible world where *C* does not occur, the effect *E* does not occur, or occurs in a substantively different way (Lewis, 2000). In the Billy and Suzie case, for example, the closest possible world in which Suzie does not throw her rock is one in which Billy does, but his rock throw will lead to a substantively different bottle shattering than the one that actually happened. At the same time, the closest possible world in which Billy does not throw is one in which Suzie throws in the same way as the actual world, resulting in the same shattering. Thus, this analysis concludes (correctly) that Suzie’s throw is a singular cause of the bottle breaking, but Billy’s throw is not. At the same time, however, there are significant concerns about this way of understanding singular causation (e.g., Kvat, 2001). For example, this theory counterintuitively implies that any factor that changes the manner of the effect’s occurrence is a singular cause of the effect’s occurrence at all. For example, putting a bandage on someone’s wound can be a cause of her death if that action delays, and so changes the manner of, her death. More significantly, this theory requires some type of distance measure over possible worlds, in order to identify the appropriate grounding for the key counterfactual. No satisfactory, substantive theory has been offered for such a measure, partly because Lewis aspired to provide a reductive account of singular causation, and so required that the distance measure not refer to any causal relations. One could instead base the distance measure partly on general causal relations, but the distance measure then becomes disposable.

The more recent difference-making theories have pursued exactly the strategy of grounding singular causation in general causal relations, informed by the actual events and (perhaps) additional information (Hall, 2007; Halpern, 2016; Halpern & Hitchcock, 2015; Halpern & Pearl, 2005b; Hitchcock, 2007a; Weslake, in press; Woodward, 2003; Woodward & Hitchcock, 2003). The shared intuition in all of these theories is that the general causal relations help to determine the relevant counterfactuals, though in a complicated manner. All of these theories have been expressed in the language of causal graphical models, so we need to have a brief detour to explain that formalism (see also Rottman, Chapter 6 in this volume). There are many introductions to graphical models (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993; also, many of the previous references in this

paragraph), and so I focus here on the high-level picture. It is important to bear in mind that the causal graphical model represents only the general causal relations; both the actual events and something more (that varies between theories) is required to get singular causation.

Causal graphical models are composed of two distinct, but related, components. The first is a directed acyclic graph that captures the qualitative causal relations (i.e., what causes what?). More specifically, we have a graph composed of nodes for the variables or events (e.g., nodes can take on different values depending on the actual state of the world), and an $A \rightarrow B$ connection just when A (the variable, or the occurrence of an event) is a general cause of B . For example, both Suzie's throw and Billy's throw are, in terms of general causal relations, causes of the target bottle breaking; they both have perfect accuracy, and so they always hit their intended targets. We can represent this qualitative causal structure as $S \rightarrow T \leftarrow B$, where the nodes correspond to Suzie's throw (S), Billy's throw (B), and hitting the target (T). Absences of arrows are informative in the causal graphical model framework: for example, the lack of an $S \rightarrow B$ edge means that whether Suzie throws does not cause whether Billy throws. The second component of a causal graphical model captures the quantitative or functional (general) causal relations. This component can take many different forms, including linear or non-linear equations, or potentially complex conditional probabilities. For example, we can use $T = S \vee B$ (where " \vee " denotes logical OR) to capture the idea that the bottle is broken if either Suzie or Billy (or both) throws. Most difference-making theories of singular causation use deterministic, quasi-logical structural equations, but deterministic systems are used principally so that the relevant counterfactuals are well-defined.² The overall framework of causal graphical models (as representations of general causal relations) is perfectly well-defined for probabilistic causal relations, whether those probabilities are due to ignorance or features of the physical situation.

The two components of a causal graphical model must be connected together in a coherent manner, typically through two assumptions. The causal Markov assumption says that every node is quantitatively independent of its non-effects (direct or indirect), conditional on its direct causes. In other words, once we know the value(s) of the direct cause(s) of node X , learning the values of nodes that are not "downstream" of X does not give us any more information about X . More generally,

the causal Markov assumption uses the qualitative graph to constrain the quantitative component. The causal faithfulness assumption (alternative, related assumptions are called Stability and Minimality) is essentially the converse of the causal Markov assumption: the only quantitative independences are those required by the causal Markov assumption. This assumption thus uses the quantitative component to constrain the qualitative graph. For example, causal faithfulness implies that any two nodes that are quantitatively independent, perhaps given knowledge of other nodes, must not be directly adjacent to one another. A key (in this context) implication of these two assumptions is that the quantitative component can be fully specified by giving the appropriate functional relation for each node in terms of its direct causes. For example, if $A \rightarrow B \rightarrow C$, then the quantitative component can be expressed as $A = f()$,³ $B = g(A)$; and $C = h(B)$.⁴ There are numerous philosophical debates about the status of these assumptions (e.g., Cartwright, 2002, 2007; Glymour, 1999; Hausman & Woodward, 1999), but we leave those aside here. In general, the motivation for these assumptions is that they encode one important way that causation can manifest in observations and data in the world.

Causal graphical models were introduced to capture general causal relations, and are particularly useful for modeling the population-level impact of manipulations, including the asymmetry of manipulation (Hausman, 1998): changing the state of a cause exogenously (i.e., from outside of the causal system) leads probabilistically to changes in its effects, but an exogenous change in an effect does not lead to changes in its causes. For example, if I change the state of a light switch, then that probabilistically leads to changes in the state of the lights; exogenously changing the state of the lights (e.g., by smashing the bulbs to guarantee that they are off) does not change the state of the switch. For convenience, I focus here on "hard" interventions that completely determine the value of the target of the intervention, though the formal framework can equally be used, with additional complications, to represent "soft" interventions that influence the target without completely controlling it (e.g., Eberhardt, 2014).

Hard interventions are easily modeled in the causal graphical model framework. To see how it works, consider the simple *Switch* \rightarrow *Lights* causal structure just mentioned. We represent a hard intervention on a target variable T by introducing a new cause I of T ; for example, we

might augment our causal structure with two specific interventions to yield $Flip \rightarrow Switch \rightarrow Lights \leftarrow Smash$. An intervention I has the special property that the value $I = yes$ (i.e., the intervention being active) completely determines the value of T and so breaks or eliminates all other edges into T . If instead $I = no$, then all of the causal relations are left intact.⁵ In our example, if $Flip = yes$, then the $Switch \rightarrow Lights$ connection is preserved. In contrast, if $Smash = yes$, then the other incoming edge to $Lights$ (i.e., $Switch \rightarrow Lights$) is broken, since the state of the lights no longer causally depends on the switch state. The asymmetry of manipulation thus emerges immediately from the (graphical) impact of interventions.

We can now return to singular causation. Recall the core counterfactual for difference-making accounts: C singularly caused E just if E would have been different if C (and perhaps other factors F, G, \dots) had been different. The causal graphical model framework can provide us with the resources to state the relevant counterfactuals more clearly. One intuitive idea is (roughly) that the key test counterfactuals arise from (1) changing C to different values by manipulation, while (2) possibly changing other variables that are *not* on a causal path from C to E in the underlying causal graph, but (3) preferring to leave these “off-path” variables at their actually occurring values if possible (for different ways of making this formally precise, see Halpern, 2016; Halpern & Pearl, 2005a, 2005b; Weslake, in press; Woodward, 2003; Woodward & Hitchcock, 2003). The “not on a path” restriction in (2) is important because we want to allow for A to be a singular cause of B even if its influence passes through the intermediate cause M , but if M is held fixed, then B 's value will not depend on A . We thus only allow ourselves to change off-path variables.

Using this overall idea, we find that Suzie is a singular cause of the bottle breaking because T would have a different value (= 0, or unbroken) if (1) S is set to a different value (= 0, or no throw) by intervention, while (2) a variable not on the $S \rightarrow T$ path, namely B , is also set to a different value (= 0, or no throw). Of course, the same analysis also shows that Billy is a singular cause, which is (by assumption) simply false. We thus see the importance of representational choices in this framework: if we want to capture the idea that Suzie's rock arrived first, then we need to explicitly represent that possibility in the causal graph, perhaps by introducing SH and BH nodes to represent Suzie or Billy's rock

hitting. When we do this, these causal graphical model-based approaches give the intuitively correct judgments.

Nonetheless, these analyses fail to capture some intuitive judgments, precisely because they focus on the actual world rather than the “normal” or “regular” world. For example, suppose that I get the influenza vaccine, but then am never subsequently exposed to the influenza virus. Intuitively, it seems incorrect to say that the vaccine is a singular cause of my not being infected, as I was never exposed in the first place. But the above-referenced theories all say that it *is* a singular cause because of the truth of the test counterfactual: if *Vaccine* were different, then *Infection* would be different (in the world in which *Exposure* occurs). Moreover, this case is formally isomorphic to ones in which these theories give the correct answer, so we have to add additional information to distinguish them. One response would be to focus on variation and covariation within a pragmatically determined, focal set of cases (Cheng & Novick, 1991), which could yield different general causal relations for this particular context. One would need a rich theory of pragmatics to fully specify this account, however.

The more common response for normative theories has been to focus only on counterfactual possibilities that are more “normal” (or closer to the “defaults,” or more “typical,” or . . .) than the actual world (different ways of capturing this idea can be found in Hall, 2007; Halpern & Hitchcock, 2015; Hitchcock & Knobe, 2009; Livengood, 2013). That is, the relevant difference-making counterfactuals for singular causation must involve changing atypical aspects to more typical ones. In the influenza case, this change means that we should not consider worlds in which *Exposure* occurs (assuming non-exposure is normal), and so the problematic counterfactual never arises. Instead, we get the intuitively correct judgment that the non-exposure is the singular cause, not the vaccine (Hall, 2007; Halpern & Hitchcock, 2015). Of course, we have to be very careful about exactly how we understand the notion of “normal” or “default” in these cases; in particular, there might be complicated multivariate patterns of normality (Livengood, 2013). Nonetheless, this adjustment can both capture our intuitions and incorporate a measure of pragmatics into our theory. On these theories, for example, oxygen is a singular cause of a fire on the International Space Station but not in my office precisely because oxygen is abnormal in space, but not in my office.

These difference-making accounts of singular causation can readily explain the cases that are difficult for the physical process theories, precisely because these accounts have no requirement that there be any consistent process, or even any process at all, connecting the singular cause with the singular effect. In particular, causation by omission is completely straightforward, since absences can clearly make a difference; absence of oxygen, for example, certainly makes a difference to one's survival. More generally, we simply need to ask, on these accounts, whether the causal factor being absent made a difference in the effect occurring. Difference-making accounts struggle, however, with overdetermination cases since those situations involve multiple factors that *could* have made a difference, but only one that actually *did* make a difference. These accounts, for example, require specific representations or default states to capture Billy not being a cause of the bottle breaking. Nonetheless, these approaches have inspired a number of experimental studies, as we will see in section "Singular Causation, Descriptively." Given the complementary strengths and weaknesses of the two types of approaches, there have been some preliminary investigations into reconciling process and difference-making theories (e.g., Woodward, 2011), though there are still many open questions about the viability of such unifications.

Worries About Both Approaches

Both types of normative theories of singular causation struggle to capture some of our intuitions, but there are more general concerns that arise for any of the currently proposed normative theories. I focus in this section on just three issues, one methodological and two substantive. The methodological worry derives from the practice of justifying a normative theory partly by demonstrating consistency with our intuitions about "important" test cases (e.g., Suzie and Billy). It is rarely explicitly stated, however, which cases should count as "important." If there are too few cases, then too many theories will be *prima facie* justifiable; every account gets the Suzie and Billy case right, for example. If we cast the net too broadly, though, then we end up with far too many cases to survey, even if we use various formal symmetries to reduce the number (Glymour et al., 2010). Instead, we need somehow to determine which test cases are truly important, and hope that there is the right number of them. No systematic position about the test cases has yet been provided, however.

The first substantive worry arises in the context of voting cases. The normative theories all yield the correct prediction for simple two-option elections, but this correctness is very fragile. As just one example (but see Livengood, 2013, for many more), suppose there are three options on the ballot and the option with the plurality (not necessarily majority) of votes wins. Suppose we have ten votes, where seven vote for option #1, two for option #2, and one for option #3. Intuitively, and on all of the normative theories (or at least, the difference-making theories; the physical process theories are less clear on these cases), the seven votes for option #1 are singular causes of that option winning. Much less intuitively, these theories say that the votes for options #2 and #3 are *also* singular causes of option #1 winning! The basic idea is that those votes being distributed as they are (in conjunction with some of the other votes) led to option #1 having the most votes (see Livengood, 2013, for proofs). Thus, they are held to be singular causes, which seems quite strange. It is unclear whether voting scenarios count as "important" test cases, or even what our exact intuitions are for complex voting cases (Glymour et al., 2010). Nonetheless, these cases reveal a significant shortcoming of the different normative theories.

The other substantive concern is arguably partially responsible for this shortcoming: namely, these normative theories do not understand singular causation as being truly contrastive. That is, they all ask "is *C* a singular cause of *E*?" rather than asking "is *C* (rather than *C**) a singular cause of *E* (rather than *E**)?" There are, however, several different lines of argument that all suggest that singular causation is fundamentally contrastive in nature (Hitchcock, 2007b; Livengood, 2013; Northcott, 2008; Schaffer, 2005). For the most part, these arguments all depend on demonstrations that whether *C* is a singular cause of *E* sometimes depends on the possible alternatives, either for *C* or *E*. For example, suppose I drink five cups of coffee and so have some muscle tremors. It seems natural to say that the coffee is a cause of the tremors, but that response is (according to proponents of contrastive accounts) based partly on our understanding that "drinking no coffee" is the natural contrast alternative. If we instead consider the contrast of drinking *eight* cups of coffee, then it is much less clear what to say. A natural response is that it was simply drinking "too much" coffee that is the singular cause, rather than any particular number of cups, but the normative theories can yield this response only if they are very careful about exactly how they represent the situation.

Relatedly, these theories include little information about the dynamics of the situation, but it seems that the (singular) cause is often thought to be the factor or factors that changed, where those changes naturally suggest exactly the contrast information that is required on these contrastive accounts (Glymour et al., 2010). Unfortunately, these normative theories typically fail to give clear answers or guidance in these types of situations.

Singular Causation, Descriptively

The different theories outlined in the previous section all aim to characterize the actual singular causation relations in the world. Our singular causation judgments presumably track this relation to some extent, but those judgments could easily diverge in particular cases, or in the content of the judgment. In general, empirical research on human singular causation judgments has been relatively independent from the normative theories; tighter integration of the two lines of research is an important challenge moving forward, as each would arguably benefit from the insights of the other. In particular, much of the empirical research blurs together judgments about multiple notions—for example, singular causation, moral responsibility, legal responsibility, and emotional reactions such as blaming—and does not carefully distinguish physical process and difference-making considerations in the experimental stimuli (with notable exceptions, of course). We thus need to be careful about exactly what conclusions we draw from particular findings of significant effects.

There are two broad types of empirical research that are informative about singular causation judgments. First, there are many experiments in which people make judgments about particular causal relations after watching a video or other perceptual sequence. Judgments such as “this collision caused that ball to move” are clearly singular in nature, and seem to emerge relatively automatically from our perceptual inputs. Many of these experiments are in the Michottean tradition, but some experiments requiring inferences about forces also involve these types of singular causation judgments. These two empirical literatures are extensively covered in White (Chapter 14 in this volume) and Wolff and Thorstad (Chapter 9 in this volume), and so I simply refer the reader to those chapters for descriptions and citations of the experiments. The key conclusion for our present purposes is that these experiments provide significant, but not unequivocal, support for physical process-style theories of

singular causation. In particular, people’s causal perceptions seem to depend on whether a continuous physical process connects the different components. At the same time, such a process seems to be sometimes inferred or imputed, arguably on the basis of difference-making features. Causal perception and force dynamics experiments have largely not systematically pitted difference-making judgments against physical process judgments to see which (if either) is driving singular causation judgments in this domain (though see Schlottmann & Shanks, 1992). At the current time, we can only conclude that some singular causation judgments seem to involve physical process considerations, though those processes need not be the sole basis of those judgments.

The second line of descriptive research is largely vignette-driven: experimental participants are provided with a story or description and are asked to judge the singular causal relations, either by identifying “what caused what” or with numeric judgments of “how much” one factor caused the target effect (see also Hilton, Chapter 32 in this volume; Lagnado & Gerstenberg, Chapter 29 in this volume). These experiments (almost) all ask for explicit conscious judgments about linguistically described situations, and the primary experimental design involves between-participant manipulations of various features of the described situation. Walsh and Sloman (2011) directly asked participants for such singular causation judgments in a number of standard cases from the philosophical literature, including ones that are structurally identical to Suzie and Billy throwing rocks. They were particularly interested in judgments of both singular causation and singular prevention of some outcome. Their results suggest that people typically use the word “cause” only when they have knowledge of some underlying physical process or mechanism, while “prevent” is often grounded in knowledge of difference-making. This suggests that “cause” and “prevent” might not be antonyms in everyday usage. At the same time, one striking feature of Walsh and Sloman (2011) is the high degree of variation in the results. Even in the seemingly straightforward Suzie/Billy case when Suzie’s rock hits the bottle first (but Billy’s would have hit it, if she had missed), 16% of participants do not agree that Suzie caused the bottle to break.⁶ More generally, there is non-trivial variation in the data for almost all of the experiments discussed throughout this section, so one must be careful not to over-interpret the results.

Lombrozo (2010) found a somewhat different pattern of singular causation judgments about typical philosophical cases, particularly those involving so-called double prevention. In a double prevention case, some factor M would normally prevent E from occurring, but a different factor C prevents M from occurring. Thus, if C occurs, then E also does. The key question is whether some particular C is a singular cause of some particular E , when this C prevents any M s from occurring. Double prevention cases clearly separate physical process and difference-making theories: there is no connection between C and E , so the former will judge C to not be a singular cause; in contrast, C made a difference for E , and so the latter will imply that C is a singular cause. Interestingly, people do not conform neatly to either theory, but rather attend to additional features of the situation. In particular, if the occurrence of C has the *function* of preventing M s so that E can occur (e.g., if someone intentionally does C to bring about E , or a machine is designed so that C leads indirectly to E), then people typically judge C to be a singular cause of E . If the C – E connection is instead accidental, then people typically judge C to not be a singular cause of E (Lombrozo, 2010). People’s singular causation judgments are influenced by the reason why C occurs, and not simply whether C occurs at all.

Other vignette-based experiments have focused less on the underlying causal structures, and more on the relationship between singular causation judgments and other types of judgments, particularly those of moral responsibility or norm violations more generally. As a concrete example, consider the widely studied “Pen Case” from Knobe and Fraser (2008). In this vignette, an administrative assistant is unable to write down an important message because the last two pens have been taken from the storage location by a professor and a staff member. The between-participant manipulation is which individuals are allowed to take a pen—the professor, the staff member, both, or neither—and participants are asked to indicate the extent to which each individual caused (or is a cause of) the inability to write down the message. Importantly, the non-social norm facts are balanced so that there is no purely physical reason to think one individual is more of a cause. Nonetheless, the standard finding is that the individual(s) who is not supposed to take a pen is judged to be more of a (singular) cause of the problem, as well as being more blameworthy. That is, singular causal judgments seem to be partially driven by moral responsibility or norm violation judgments (e.g., Alicke, 1992; Alicke,

Rose, & Bloom, 2011; Hitchcock & Knobe, 2009; Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015), though general causal judgments are interestingly less influenced by such considerations (Danks, Rose, & Machery, 2014). At the same time, the influence also seems to go in the other direction: for example, causal judgments can influence moral culpability judgments (Cushman, Knobe, & Sinnott-Armstrong, 2008).

At a high level, the results of these types of vignette-based experiments are largely consonant with the more sophisticated difference-making accounts of singular causation. That is, people’s singular causal judgments seem to be sensitive to the truth of particular focal counterfactuals that can be derived from (a) causal graphical model representations of the general causal relations, and (b) facts about the specific situation, including defaults or “normal” values. We have independent grounds for thinking that causal graphical models provide a good model of human causal knowledge (e.g., Danks, 2014; Holyoak & Cheng, 2011; Rottman, Chapter 6 in this volume), and so much of the focus of this research has been on the details of (b), and particularly on the role of the norms—statistical, social, conventional, or moral—present in the actual situation.

Early work in this area focused on the influence of prescriptive norms that say what one ought to do, or how something ought to function. For example, someone acting illegally is judged to be more of a (singular) cause of some bad outcome than an individual acting legally (Alicke, 1992). The relevant norms need not be legal ones, though, as shown by the Pen Case: the singular causation judgments in that case track violations of the prescriptive norm, even though that norm is grounded in social and institutional facts, rather than legal ones. More generally, a number of different experiments have shown that a violation of a prescriptive norm is consistently judged to be more of a singular cause than the exact same action when no prescriptive norm is being violated (e.g., Alicke, 1992; Alicke et al., 2011; Cushman et al., 2008; Hitchcock & Knobe, 2009; Knobe & Fraser, 2008). There is even suggestive evidence that singular causation judgments are influenced by violations of typicality norms that indicate what is statistically normal in a population (Hitchcock & Knobe, 2009; though see Sytsma, Livengood, & Rose, 2012). Moreover, these singular causal judgments are sensitive not just to whether the action violated a norm, but also whether *other* relevant actions or events violated

a norm; in particular, the causal responsibility of a norm-violating event can be mitigated when another norm-violating event “supersedes” the former (Kominsky et al., 2015).

One challenge in interpreting these experiments is that there is often no clear understanding of what additional information is carried by some prescriptive or typicality norm, or by a violation of that norm. Suppose, for example, that the Pen Case norm is that professors are not supposed to take pens (and so the professor is judged to be more of a singular cause). Given only knowledge of the existence of this norm, one could potentially, but not necessarily, also infer that probably (a) in the past, there have been problems when professors took pens; (b) in the past, there have not been problems when staff members took pens; (c) professors only take pens when they have a reason that overrides the norm; (d) professors usually do not take pens; (e) staff members take pens more often than professors; (f) common knowledge of any of (a)–(e) within the department; and perhaps many other implications. That is, if we learn that some action is a norm violation, we often learn more than just that there is a norm; we also potentially learn many additional facts about the relevant causal structures, statistics, and possible actions. It is quite clear that norm violation information influences singular causation judgments, but the how and why of that influence is largely unknown. Further research carefully disentangling these different pathways is a significant open research problem.

Social psychology research on attribution theory (e.g., Heider, 1958; Hilton, Chapter 32 in this volume; Kelley, 1973; Nisbett & Ross, 1991) provides an additional set of empirical results that have arguably been underutilized in the study of singular causation judgments. Attribution theory examines the (causal) explanations that people provide to explain their own or others’ behaviors, particularly focusing on whether those explanations appeal to factors that are internal or external to the agent. For example, if I am late to a meeting, is that explained in terms of some internal disposition (“David always loses track of the time”) or external circumstances (“The bus that David takes to campus arrived late”)? The most famous result—the so-called fundamental attribution error (Jones & Harris, 1967)—was that people (or at least, Western-educated undergraduate students) tend to emphasize internal factors when explaining others’ actions, but external factors when explaining their own. As with almost all “classic” findings, the story is significantly more complicated

than the usual presentations of the fundamental attribution error (Malle, Knobe, & Nelson, 2007; Norenzayan, Choi, & Nisbett, 2002). Nonetheless, this overall area of social psychological research is clearly relevant, as people are making singular causation judgments, perhaps implicitly, in the course of constructing particular causal explanations. There has been some crossover between these literatures, but not yet a systematic integration of the two (see also Hilton, Chapter 32 in this volume).

For all of these experimental results, it is also important to bear in mind the potential limits of vignette-based research. There is a long history of experiments in judgment and decision-making demonstrating that people behave differently if information is presented as a (textual) story—“learning from description”—rather than as cases or other less-linguistic stimuli—“learning from experience” (Barbey & Sloman, 2007; Erev et al., 2010; Gigerenzer & Hoffrage, 1995; Hau, Pleskac, Kiefer, & Hertwig, 2008; Hertwig & Erev, 2009). The general pattern of findings about these two modes of learning is that estimation and reasoning are more accurate given learning from experience (rather than learning from description), particularly in domains such as contingency learning and choice under uncertainty. In particular, people’s judgments seem to be less subject to factors that are seemingly irrelevant to those tasks, such as salience or representativeness of particular stimuli. The exact mechanisms underlying these differences are currently an open research question, so it is unclear whether similar behavior should arise in singular causal judgment. At the same time, people’s singular causation judgments given more naturalistic stimuli could plausibly be quite different from the results reported here. For example, normative considerations could conceivably play less of a role if people learn from experience rather than a vignette (see also Danks et al., 2014).

Finally, there is an interesting feature that emerges from all of these lines of empirical research, and that calls into question whether the modifier “singular” is appropriate. The use of that modifier suggests that the judgment is specific to this particular, unique situation, rather than applying more generally. However, we consistently find that people use “singular” causation judgments, whether based in perception, vignettes, or prior knowledge, to make inferences about other cases. That is, people seem to regard judgments of singular causation as “portable,” in the sense that they are informative about, and can carry over to, novel situations

(Danks, 2013; Hitchcock, 2012; Lombrozo, 2010). As just one example, perceptual judgments about collisions between balls—that is, paradigmatic cases of causal perception—provide generalizable information about future collisions, such as the relative weights of the balls (White, 2009). In some ways, the exportability of singular causation judgments is unsurprising: there would be little reason to make such judgments if they were completely uninformative about future situations. Nonetheless, this feature has been largely ignored in normative theories of singular causation (though see Hitchcock, 2012). Despite the name, singular causation judgments are not fully unique and particularized, but rather seem to be connected closely with our abilities to act, predict, and explain in future, novel situations.

Open Problems and Challenges

There has been substantial research on singular causation—both psychologically and philosophically—but many open questions remain. Perhaps the most obvious open challenge is that, despite the experimental results discussed in the section “Singular Causation, Descriptively,” the relevant empirical phenomena and underlying cognitive processes have been only partially characterized. Some situational and psychological influences on singular causation judgments have been identified, but it is unlikely that these form a complete set. Moreover, very little is currently known about the cognitive processes by which these factors influence those judgments. For example, it seems quite likely that blame judgments (or other morally negative appraisals) influence causal judgments, but there are many possible routes by which they could come to have such impact. Consider three possible, not mutually exclusive, mechanisms inspired by different theoretical proposals (for singular causation) currently in the literature: (1) morally negative appraisals (or other judgments of norm violation) trigger counterfactual thinking, which prompts particular singular causation judgments (Knobe, 2009; Kominsky et al., 2015); (2) people want (perhaps unconsciously) to justify their negative reactions, and so they judge the targets of those reactions as more causal since one can only blame or criticize something that was a cause (Alicke, 1992); and (3) morally wrong behavior or other norm-violating factors are those one most wants to change, and singular causation judgments carry information about which factors are “good” candidates

for intervention (Hitchcock & Knobe, 2009). These three possible mechanisms presumably can be experimentally distinguished if we expand our methods beyond vignette studies to include, for example, eye-tracking and reaction-time studies. The relevant experiments have not yet been performed, however, and so the underlying cognitive mechanisms remain an open question. On the normative, philosophical side, theories of singular causation remain an active topic of research interest, with a particular emphasis on developing accounts that are better grounded in both formal theories and established intuitions.

Another significant set of open problems centers on the relationships between singular causation judgments and causal explanations (see also Lombrozo & Vasilyeva, Chapter 22 in this volume). In many cases, singular causal judgments are made partly to provide key premises in a causal explanation of some particular event. For example, if I am trying to explain why there is broken glass on the ground, I might appeal to the fact that “Suzie’s rock caused the bottle to break.” That is, the singular causation judgments do not solely describe the world, but also play an important role in our causal explanations. Thus, to the extent that our causal explanations are not simply lists of possibly relevant facts, we should expect our singular causation judgments to potentially bear some hallmarks of that function. Lombrozo (2010) explored this connection between explanation and singular causation judgments, and found that those latter judgments did seem to be sensitive to their subsequent use in causal explanations. In particular, just as causal explanations generalize to similar situations, singular causation judgments were shaped by their generalizability to future cases (see also Ahn & Bailenson, 1996). At the same time, a singular causation judgment is obviously not the same as a causal explanation, and there appear to be some systematic divergences between judgments and explanations (Livengood & Machery, 2007). The exact nature of those differences is a substantial open problem.

A final significant challenge is the lurking possibility of pluralism about singular causation, either psychologically or philosophically. An implicit assumption of this whole chapter has been that there is some single relation—whether in the world or in our minds—that corresponds to singular causation. The concern is that this assumption might

be false: perhaps singular causation corresponds to many different relations depending on domain, background knowledge, and so forth. That is, perhaps there is no single description that picks out the singular causes in either the world or our judgments. A number of philosophers (e.g., Hall, 2004) have suggested that there are two different types of singular causation (either relations or judgments): one type based on physical processes, and one based on difference-making. Of course, in the actual world, these two types of singular causation typically proceed in unison: if there is a physical process connecting C with E , then C will make a difference in E , and vice versa. Nonetheless, these are distinct relations, and the pluralists argue that there is no single relation that corresponds with singular causation, again either in the world or in our judgments. These arguments leave open the possibility that some future account will successfully unify these different approaches; Wolff (2014), for example, argues that force dynamics models provide such a unification (see also Wolff and Thorstad, Chapter 9 in this volume). Despite the implicit assumption in most of this chapter of monism about singular causation, the precise taxonomy of singular causation—one versus many—remains a significant open challenge.

It is clear that many different cognitive operations depend critically on judgments of, and reasoning about, singular causation. An understanding of general causal structure does not suffice for causal explanations of particular events, or rich counterfactual reasoning about a particular case, or assignment of blame for particular outcomes. Instead, we need to use additional information, whether about physical processes connecting parts of the causal structure, or about which factors (counterfactually) made a difference about the target event. This area of causal cognition is somewhat unusual, as our normative, philosophical understanding is arguably more advanced than our empirical, psychological understanding. Increased interaction between the two approaches can thus only help to advance our knowledge of this area.

Notes

1. Technically, Suzie's rock is interacting with other causal processes (e.g., air molecules) throughout its flight, as a tiny amount of its mass-energy is being transferred to those molecules. I ignore this complication.
2. There are theories of counterfactuals for inherently probabilistic situations (e.g., "if I had worn a blue shirt, then that radioactive atom would not have decayed"), but little

agreement about even what intuitions or phenomena should be captured by such theories.

3. That is, A 's value is set by some exogenous factors outside of the causal system.
4. Note that "=" in these equations includes information about causal order; $B = h^{-1}(C)$ might hold mathematically (when h is invertible), but it fails to represent the causal relations.
5. In terms of the corresponding structural equations, $I = yes$ means that the equation for T changes from a function of T 's graphical parents to simply $T = t_p$ where t_p is the T -outcome of the intervention.
6. See the Mechanism Complete condition of Experiment 5 in Walsh and Sloman (2011). They use a slightly different cover story and swap names, but it is structurally identical to the Suzie/Billy case that has been a running example in this chapter.

References

- Ahn, W.-K., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology*, 31, 82–123.
- Alicke, M. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 36, 368–378.
- Alicke, M., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy*, 108(12), 670–696.
- Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30, 241–297.
- Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward. *The British Journal for the Philosophy of Science*, 53, 411–453.
- Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge: Cambridge University Press.
- Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, 40, 83–120.
- Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, 108, 281–289.
- Danks, D. (2013). Functions and cognitive bases for the concept of actual causation. *Erkenntnis*, 78, 111–128.
- Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. Cambridge, MA: MIT Press.
- Danks, D., Rose, D., & Machery, E. (2014). Demoralizing causation. *Philosophical Studies*, 171(2), 251–277.
- Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59, 195–216.
- Dowe, P. (2001). A counterfactual theory of prevention and "causation" by omission. *Australasian Journal of Philosophy*, 79, 216–226.
- Dowe, P. (2007). *Physical causation*. Cambridge: Cambridge University Press.
- Eberhardt, F. (2014). Direct causes and the trouble with soft interventions. *Erkenntnis*, 79(4), 755–777.
- Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., et al. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, 23(1), 15–47.

- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684–704.
- Glymour, C. (1999). Rabbit hunting. *Synthese*, 121, 55–78.
- 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. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 225–276). Cambridge, MA: MIT Press.
- Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, 132(1), 109–136.
- Halpern, J. Y. (2016). *Actual causality*. MIT Press.
- Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *British Journal for the Philosophy of Science*, 66(2), 413–457. doi:10.1093/bjps/axt050
- Halpern, J. Y., & Pearl, J. (2005a). Causes and explanations: A structural-model approach, Part I: Causes. *The British Journal for the Philosophy of Science*, 56, 853–887.
- Halpern, J. Y., & Pearl, J. (2005b). Causes and explanations: A structural-model approach, Part II: Explanations. *The British Journal for the Philosophy of Science*, 56, 889–911.
- Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, 21(5), 493–518.
- Hausman, D. M. (1998). *Causal asymmetries*. Cambridge: Cambridge University Press.
- Hausman, D. M., & Woodward, J. (1999). Independence, invariance and the causal Markov condition. *The British Journal for the Philosophy of Science*, 50, 521–583.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley and Sons.
- Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13, 517–523.
- Hitchcock, C. (2007a). Prevention, preemption, and the principle of sufficient reason. *The Philosophical Review*, 116(4), 495–532.
- Hitchcock, C. (2007b). Three concepts of causation. *Philosophy Compass*, 2–3, 508–516.
- Hitchcock, C. (2012). Portable causal dependence: A tale of consilience. *Philosophy of Science*, 79(5), 942–951.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 106(11), 587–612.
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62, 135–163.
- Hume, D. (1748/2001). *An enquiry concerning human understanding*. Oxford: Clarendon.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3(1), 1–24.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28, 107–128.
- Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific Explanation* (pp. 410–505). Minneapolis: University of Minnesota Press.
- Knobe, J. (2009). Folk judgments of causation. *Studies in the History and Philosophy of Science*, 40, 238–242.
- Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (pp. 441–448). Cambridge, MA: MIT Press.
- Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, 137, 196–209.
- Kvart, I. (2001). Counterexamples to Lewis' "Causation as influence." *Australasian Journal of Philosophy*, 79, 411–423.
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556–567.
- Lewis, D. (2000). Causation as influence. *Journal of Philosophy*, 97, 182–197.
- Livengood, J. (2013). Actual causation and simple voting scenarios. *Nous*, 47(2), 316–345.
- Livengood, J., & Machery, E. (2007). The folk probably don't think what you think they think: Experiments on causation by absence. *Midwest Studies in Philosophy*, XXXI, 107–127.
- Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, 61, 303–332.
- Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 93(4), 491–514.
- McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies*, 123, 125–148.
- Menzies, P. (1996). Probabilistic causation and the pre-emption problem. *Mind*, 105, 85–117.
- Nisbett, R. E., & Ross, L. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.
- Norenzayan, A., Choi, I., & Nisbett, R. E. (2002). Cultural similarities and differences in social inference: Evidence from behavioral predictions and lay theories of behavior. *Personality and Social Psychology Bulletin*, 28, 109–120.
- Northcott, R. (2008). Causation and contrast classes. *Philosophical Studies*, 139, 111–123.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.
- Salmon, W. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64, 461–477.
- Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
- Schaffer, J. (2000). Causation by disconnection. *Philosophy of Science*, 67, 285–300.
- Schaffer, J. (2001). Physical causation. *British Journal for the Philosophy of Science*, 52, 809–813.
- Schaffer, J. (2005). Contrastive causation. *Philosophical Review*, 114, 297–328.
- Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology*, 44A, 321–342.
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search* (1st ed.). Berlin: Springer.
- Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(4), 814–820.
- Walsh, C. R., & Sloman, S. A. (2011). The meaning of cause and prevent: The role of causal mechanism. *Mind and Language*, 26(1), 21–52.

- Weslake, B. (in press). A partial theory of actual causation. *The British Journal for the Philosophy of Science*.
- White, P. A. (2009). Perception of forces exerted by objects in collision events. *Psychological Review*, 116(3), 580–601.
- Wolff, P. (2014). Causal pluralism and force dynamics. In B. Copley & F. Martin (Eds.), *Causation in grammatical structures* (pp. 100–119). Oxford: Oxford University Press.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
- Woodward, J. (2011). Mechanisms revisited. *Synthese*, 183, 409–427.
- Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, Part I: A counterfactual account. *Noûs*, 37(1), 1–24.

