Published in “Counterfactuals and Causal Reasoning,” in E. Siegal and N. Boneh (eds.), *Perspectives on Causation*, Springer, 2020: 463–82.

COUNTERFACTUALS AND CAUSAL REASONING

Boris Kment

Abstract. Counterfactual conditionals are used extensively in causal reasoning. This observation has motivated a philosophical tradition that aims to provide a counterfactual analysis of causation. However, such analyses have come under pressure from a proliferation of counterexamples and from evidence that suggests that the truth-conditions of counterfactuals are themselves causal. I offer an alternative account of the role of counterfactuals in causal thought that is consistent with these data: counterfactuals are used in a common method of causal reasoning related to John Stuart Mill’s method of difference. The method uses background beliefs about causal relationships, history, and the natural laws to establish a new causal claim. Counterfactuals serve as a convenient tool for stating certain intermediate conclusions in this reasoning procedure, and that is part of what makes counterfactuals useful. This account yields a functional explanation of why our language contains a construction with the truth-conditions of counterfactuals.

1. Causation and counterfactuals: the order of explanation

Counterfactual thought is an important element of our cognitive lives. In making practical decision, we are often led to ask what would happen if we were to carry out a certain action, and we frequently support causal claims by showing that the putative effect depends counterfactually on the supposed cause. It therefore does not come as a surprise that scholars in many disciplines—from philosophy to cognitive and social psychology to computer science to linguistics—have shown a keen interest in understanding counterfactuals.[1]

One point of contention is whether causal notions should figure in a semantic account of counterfactuals. A number of philosophers, motivated by examples like those described in section 2 below, have favored such causal theories of counterfactuals. However, this approach stands opposed to a prominent philosophical tradition, going back at least to David Hume and most prominently defended by David Lewis,[2] that aims to give a reductive analysis of causation in counterfactual terms. The two views advocate for opposite directions of analysis and are consequently mutually exclusive—combining them would lead to circularity.

[1] See Roese et al. (2014) and Hoerl et al. (2011) for contributions by psychologists, Pearl (2009: ch. 7) for discussion by a computer scientist and philosopher, and the references in the rest of this paper for further literature on counterfactuals.
[2] Hume (1995: 87), Lewis (1986a, 2004a).

The goal of this paper is to provide further support for a causal account of counterfactuals, by showing that it can explain the phenomenon that motivates counterfactual analyses of causation as convincingly as these analyses themselves can. The phenomenon under consideration is our pervasive tendency to use counterfactuals to support causal claims. A counterfactual account of causation provides a straightforward explanation of this datum: causal relationships *consist* (at least partly) in certain patterns of counterfactual dependencies, and to ask whether *X* is a cause of *Y* is therefore (at least in part) to ask whether certain counterfactuals hold. Nevertheless, counterfactual analyses face considerable obstacles, which I will review in section 2. That will provide initial motivation to look for an alternative explanation of the role of counterfactuals in causal reasoning and will set the stage for my own account.[3] I will offer a somewhat idealized and simplified rational reconstruction of our use of counterfactuals in ordinary-life causal reasoning, focusing on deterministic contexts in section 3 and on indeterministic ones in section 4. On my account, the procedure uses certain background beliefs about causal relationships, natural laws, and history, to establish a new claim about relationships of (actual token) causation. Counterfactuals serve as a convenient tool for stating certain intermediate conclusions in this reasoning procedure, and that is one of the reasons why we have a counterfactual construction. This account yields a good functional explanation of why our language contains a construction with the truth-conditions of counterfactual conditionals: they are exactly the truth-conditions that a construction needs to have to adequately serve the purpose of stating the intermediate conclusion in the reasoning practice I have described. As we will see, these truth-conditions involve the concept of causation.[4,5]

2. The counterfactual analysis of causation and the causal account of counterfactuals

Y counterfactually depends on *X* just in case *Y* would not have existed (obtained, occurred) if *X* had not existed (obtained, occurred). Counterfactual accounts of causation aim to analyze causation in terms of counterfactual dependence. I will discuss three of the main challenges confronting this approach. *Firstly*, for the counterfactual account to be tenable, there must be necessary and sufficient condition for causation that can be stated in counterfactual terms. However, it is very hard to find such conditions, even for causation under determinism. *Secondly*, even if this problem could be solved for deterministic

[3] For an interesting alternative explanation of the connection between causation and counterfactuals, see Maudlin (2004).

[4] I do not claim that that is the only function of counterfactual conditionals. They clearly also serve other purposes, e.g. in making practical decisions. (See, e.g., Stalnaker (1981), Gibbard and Harper (1981), Lewis (1981), and Joyce (1999). For arguments against counterfactual decision theory, see Ahmed (2014).) See Edgington (2003) and Bennett (2003) for discussions of further uses of counterfactuals.

[5] For a much fuller development of the view proposed in this paper, see Kment 2010, and in particular Kment 2014: Chs. 10–12. Also see Kment 2015: Sct. 5.

causation, it would be hard to extend the account to probabilistic causation (which presents its own set of challenges). *Thirdly*, one may not find the main motivation for pursuing a counterfactual analysis of causation particularly compelling.

Start with the first point: the difficulty of formulating necessary and sufficient conditions for causation. The main problem is that simple counterfactual dependence between distinct matters of particular fact is not a necessary condition for causation.[6] There are two types of cases that are commonly used to show this.

Over-determination. Fred's rock and Susie's rock simultaneously hit the window, each causing sufficient damage to shatter it. Fred's throw and Susie's throw are both causes of the shattering (the shattering is causally overdetermined), but the shattering does not counterfactually depend on either throw. If one of the throws had not occurred, the other would still have broken the window.

Preemption. Susie and Fred are both getting ready to throw a rock at the window. Susie throws hers first and shatters the window, thereby preempting Fred's plan. Susie's throw is a cause of the breaking of the window, but the breaking does not counterfactually depend on her throw. If she had not thrown her rock, Fred would have thrown his, which would have shattered the window.

As these examples show, effects need not counterfactually depend on their causes. Consequently, if causation is to be analyzed in counterfactual terms at all, it cannot be analyzed as simple counterfactual dependence, but at best as some *complex pattern* of counterfactual dependencies (possibly combined with other conditions). There have been numerous attempts to provide such an analysis in a way that gets the extension of causation right, but in my opinion they met with only limited success.[7,8]

We face additional obstacles when trying to give a counterfactual analysis of probabilistic causation. If indeterminism is pervasive, so that it is almost always a matter of chance what happens, then it is almost never true that the effect would not have happened if the cause had not happened. The effect still *might* have happened (though it might or would have been less likely). For that reason alone, it seems hopeless to try to define

---

[6] By "matters of particular fact" I mean, roughly speaking, facts about the goings-on in specific space-time locations.

[7] For some of the strategies for dealing with overdetermination and preemption problems within the framework of the counterfactual account, see Lewis (1986a), Menzies (1989), McDermott (1995), Ramachandran (1997), Lewis (2004a), Yablo (2004). For an overview, see Paul and Hall (2013). More recently, philosophers using the framework of causal models have proposed a number of other treatments of overdetermination and preemption problems. See, e.g., Hitchcock (2001), Woodward (2003), Halpern and Pearl (2005), and Hall (2007). Other philosophers have tried to address the problems for counterfactual accounts in part by arguing that there are several concepts of causation, and that the problems arise from choosing the wrong notion as the target for a counterfactual analysis (Hall 2004; for a reply, see Kment 2014: Sct. 9.1.2, in particular p. 225 n.1, Sct. 10.4.1).

[8] There are also cases that seem to show that counterfactual dependence between distinct matters of particular fact is not sufficient for causation (see Bennett 1984; also Kment 2010: 84–5, 2014: 248–9). These examples have been discussed less extensively.

probabilistic causation in terms of the counterfactual dependence of the effect on the cause. Instead, philosophers have tried to define probabilistic causation in terms of the counterfactual dependence of the *chance* of the effect on the cause. The most popular version of this account starts from the idea that causes are probability raisers, a thought that can be refined in various ways. (For a classic statement of this view, see Lewis (1986a: postscript B). Also see Menzies (1989).)

Unfortunately for this approach, whether *C* is a cause of *E* cannot be read off the way in which *E*'s chance counterfactually depends on *C*. Consider an example due to Jonathan Schaffer (Schaffer 2000). Merlin casts a spell to turn the prince and the king into frogs at midnight, and Morgana casts a spell to turn the prince and the queen into frogs at midnight. Once one of these spells has been cast, its chance of success remains constant at 50% until midnight. Since the results of the two spells are stochastically independent, the two spells result in a chance of 75% that the prince will become a frog at midnight. At midnight the prince is transformed along with king, while the queen is not. The result proves that Merlin's spell worked, while Morgana's was ineffective. So, Merlin's spell is a cause of the prince's transmutation while Morgana's is not. Note, however, that each spell raised the probability of the prince's transmutation by the same amount, from 50% to 75%. The example shows that the way in which a factor influences the chance of *X* need not determine whether it makes a causal contribution to the occurrence of X.

It seems plausible enough that we can use counterfactuals about chances to support claims about the causes *of these chances*. If we know that *E* would not have had chance *p* at time *t* if *A* had not obtained, then we have reasons for concluding that *A* is a cause of the fact that *E* had chance *p* at *t*. But, given that indeterministic causation of *other* effects (i.e., effects that are not facts about chances) is not merely a matter of influencing chances, it is not obvious how to extend the account to such instances of causation. However, such an extension would be needed to obtain a unified counterfactual account that covers all cases of causation.

Not only do counterfactual accounts of causation face these obstacles, but the considerations motivating the analysis seem very resistible. The analysis receives its greatest support from its ability to explain our tendency to infer causal claims from counterfactual connections. However, we have an equally strong tendency to draw the contrapositive inference from the *absence* of a causal connection to counterfactual *in*dependence. Moreover, just as beliefs about causal relationships are often guided by counterfactual judgments, so counterfactual judgments are frequently based on prior causal beliefs. We could appeal to the former phenomenon to support an analysis of causation in counterfactual terms, but we could equally well appeal to the latter phenomenon to motivate a causal account of counterfactuals.

Before considering a case in which our counterfactual judgments are guided by causal beliefs, we need a simplified working account of the truth-conditions of counterfactual conditionals. I will endorse the standard account of their truth-conditions in terms of comparative closeness or overall similarity between possible worlds. To simplify somewhat, ⌜If it had been that *P*, then it would have been that *Q*⌝ is true just in case *Q* is true at the *P*-worlds closest to the actual world.[9] It is well-known that the operative standards of overall similarity between worlds must differ from those underlying our offhand similarity judgments. It sounds true to say "If Nixon had pressed the button, then there would have been a devastating nuclear war." That means that the closest worlds where Nixon presses the button are those where the earth is devastated, not those where the signal dies in the wire on its way to the launch pad, even though offhand we would judge the latter worlds to be more similar to actuality than the former.[10]

What are the standards of inter-world similarity that matter to the truth-conditions of counterfactuals? Suppose that Susie throws a rock at the window and the window shatters, and ask yourself whether the window would still have shattered if Susie had not thrown the rock. Roughly speaking, the closest possible worlds where Susie doesn't throw the rock are those that are just like the actual world at the time of the rock throw, except that Susie does not throw her rock. After that, the world evolves in accordance with the natural laws of the actual world. If the window breaks in that world, then we can express this by saying that the window would still have shattered if Susie had not thrown her rock. Otherwise, it is true to say that the window would not have shattered if Susie had not thrown her rock, i.e. that the shattering depends counterfactually on Susie's throw. We can generalize from this example. Let *A* and *E* be matters of particular fact that actually obtain, with *A* obtaining at time t_A and *E* obtaining at some later time:

(1) Under determinism, *E* counterfactually depends on *A* just in case, at every possible world that is like actuality at t_A except that *A* does not obtain and that conforms to the actual laws of nature after t_A, *E* fails to obtain.[11,12]

---

[9] Throughout this paper, I will make the simplifying assumption that the "limit assumption" is true, i.e. that for any antecedent, there is a set of antecedent-worlds that are equally close to actuality and closer than any antecedent-worlds not in the set. Although this assumption is likely to be false, the simplification is harmless. For, there are well-known ways of doing without the limit assumption (Lewis 1973) and they could easily be applied (with some loss of simplicity) to the discussion in this paper.

[10] See Bennett (1974), Fine (1975), Lewis (1973: 76), (1986b). The Nixon case is a variant of Fine's example.

[11] This is simplified in a number of ways. For example, there is, strictly speaking, no possible world where *A* fails to obtain but which matches actuality in *all* facts about t_A other than *A*. (For, some of the facts about t_A other than *A* necessitate *A*, e.g. the fact that *A* and *B* both obtain, where *B* is some other fact about t_A.) A more precise description of the closest worlds where *A* fails to obtain would say that these worlds *maximize* match in facts about t_A other than *A*. That is to say, of all the worlds where *A* fails to obtain, the closest ones (other things being equal) are those that come closest to matching actuality in all the facts about t_A other than *A*. Of course, this account is still simplified. For a fuller and more precise account, see Kment 2006, 2014: Chs. 8–9.

[12] Counterfactuals are notoriously context-dependent (Quine (1950), Lewis (1973), (1986b)); different standards of similarity are relevant to their truth-conditions in different contexts. However, like David Lewis (1986b), I believe (Kment 2006: 262–3, 2014: 44–46) that there is a specific standard of closeness that serves as our default—we use this standard in interpreting and evaluating counterfactuals unless our presumption in its favor is canceled by distinctive features of the context. Lewis's account of causation analyzes causation in terms of this default standard of closeness. (1) and (2) describe the conditions for counterfactual

Under determinism, the state of an antecedent-world at t_A and the actual laws of nature together determine the entire rest of history. The same is not generally true under indeterminism. Two antecedent-worlds might both be like actuality at t_A except for the fact that *A* does not obtain, and they might both conform to the actual laws thereafter, and yet they may differ in the outcomes of some post-t_A random processes. That raises the question whether similarity to the actual world in the outcomes of post-t_A chance processes is an additional criterion of closeness to actuality.

The answer is a qualified 'yes.' Some post-antecedent similarities matter to the closeness ordering, others do not. Consider a variant of an example due to Dorothy Edgington (2003, 2011). You are about to watch an indeterministic lottery draw on television when someone offers to sell you ticket number 17. You decline. As luck would have it, ticket number 17 wins. It seems true to say, "If you had bought the ticket, you would have won." But that presupposes that the following is true:

> If you had bought ticket number 17, that ticket would still have won.

Now suppose that the lottery company has two qualitatively indistinguishable lottery machines that give the same chance to every possible outcome. They used machine A in the draw but could have used machine B instead. Consider:

> If a different machine had been used, 17 would still have won.

That seems false. If a different machine had been used, then 17 might still have won, or some other number might have won. It is not true that 17 would still have won. In the first case, we hold the outcome of the lottery draw fixed, in the second we do not. It seems very plausible that this difference is due to underlying causal judgments. Your decision about whether to buy the ticket is not causally connected to the outcome of the draw (or so we believe). That is why the outcome can be held fixed when we are thinking about what would have happened if you had made a different decision. By contrast, the use of a particular lottery machine is part of the causal history of the outcome. That is why the outcome of the draw cannot be held fixed in the second case. In these examples, we are drawing on prior causal judgments to decide whether certain facts can be held fixed—i.e., whether they would still have obtained if the antecedent had been true, or in other words, whether they are counterfactually independent of the antecedent.[13]

The upshot is that, when we think about what the world would be like if *A* had not obtained, we are holding fixed just those post-antecedent matters that are not causally connected to *A* in the actual world. For the indeterministic case, therefore, we can give the

---

dependence under the default standard. Moreover, the method of evaluating causal claims in the light of counterfactual dependencies that I will discuss in this paper employs the default standard as well.

[13] Examples like this are sometimes called "Morgenbesser cases," in honor of Sydney Morgenbesser, who was among the philosophers who discovered them (although Morgenbesser did not publish the result). Examples similar to the one described are discussed in Adams (1975: ch. IV, sct. 8, in particular pp. 132–3.), Tichý (1976), Slote (1978), Bennett (2003), Edgington (2003, 2011), Schaffer (2004), and Kment 2006: Scts. 3–4, 2014: Chs. 8–9, in particular Scts. 8.3–8.4.

following, somewhat simplified account of the conditions for counterfactual dependence. Suppose that *A* and *E* are actual matters of particular fact, with *A* actually obtaining at t_A and *E* actually obtaining at some later time.

(2) Under indeterminism, *E* counterfactually depends on *A* just in case *E* does not obtain at any world *w* that meets the following conditions:

(a) *A* fails to obtain at *w*,

(b) *w* is otherwise like actuality at t_A,

(c) *w* matches actuality after t_A in all matters of particular fact that are not actually caused by *A*, and

(d) *w* conforms to the actual laws after t_A.

It seems that the conditions for counterfactual dependence are themselves causal.[14] If that is true, then causation cannot be analyzed in terms of counterfactual dependence without circularity.

3. Counterfactual dependence and deterministic causation

3.1 The determination idea

I will argue that our practice of using counterfactuals to evaluate causal claims rests on an assumption I will call the "determination idea." Separate versions of this idea apply to deterministic and to indeterministic contexts.[15] The deterministic version will be considered in this section and the indeterministic version in section 4.

The deterministic version of the determination idea ("D/d," for short) runs as follows:

(D/d) Under determinism, the causes of *E* together nomically determine *E*.

That is to say: *E* obtains at every possible world where all the actual causes of *E* obtain and which conform to the actual laws of nature.

Some clarification is in order concerning the notion of cause used in (D/d). We can distinguish two ways in which *x* can cause *y*.[16] On the one hand, *x* might be part of what

[14] See Mårtensson (1999), Edgington (2003, 2011), Bennett (2003: ch. 15), Hiddleston (2005), and Wasserman (2006) for causal analyses of counterfactuals motivated by examples like the above lottery case, and see Kment 2006, 2014: Chs. 8–9 for an analysis in terms of (causal and non-causal) explanation. Veltman (2005) and Schulz (2011) take a similar line. Also see Pearl (2009, ch. 7), who uses the framework of causal models to give a causal account of counterfactuals and of what is held fixed in counterfactual reasoning. For an early causal theory of counterfactuals, see Jackson (1977).

[15] By "determinism" I mean the thesis that the state of the universe at any given moment and the laws of nature together determine all of history: any possible world that matches actuality at one time and that conforms to the actual laws of nature matches actuality at all times.

[16] For more on this distinction, see Ned Hall's discussion in his (2004) and David Lewis's in his (2004b). I don't agree with their thesis that we need a counterfactual account of causation (or of one notion of causation) to accommodate the thought that omissions are causes. I think that our belief in omissions as causes is closely

produced *y* (the stone throw produced the breaking of the window, the poisoning caused the patient's death, etc.). On the other hand, *x* might cause *y* without being among the producers of *y*. These non-producing factors include omissions, such as the absence of various kinds of possible interference with the causal processes that produced *y*.[17] They also include factors (so-called "double preventers") that prevent such interferences and thereby cause their absence, as well as the producing and non-producing causes of such double preventers. For example, the fact that the would-be assassin failed to kill the president on the eve of her speech forms part of the causal history of the speech, as does the action of the police agent who arrested the assassin before he could strike. (It is partly because of the action of the police agent and the absence of assassins that the president holds the speech the next day.) But neither the police agent's action nor the absence of assassins is among the factors that produced the speech. The notion of cause used in (D/d) is to be understood in a broad sense, as including not only the producers of *E*, but also *E*'s various non-productive causes.[18]

For the purpose of illustration, assume that determinism is true and suppose that Susie throws a rock at a window and breaks it. Consider all the causes of the window shattering, *including omissions*. These causes include Susie's throw, the position and molecular structure of the window, etc. They also include the absence of any factors that could interfere with the shattering, such as obstacles in the path of the flying rock, strong winds that could blow the rock off its path, bystanders trying to catch the rock, and so forth. (D/d) tells us that, if you complete this list of causes in the right way, then you get a set of factors that nomically determines the breaking of the window.

Note that the determination idea merely states a *necessary* condition for causation; it does not state a sufficient condition. That is to say, it is true of a set of factors that it contains all and only the causes of *E* only if that set nomically determines *E*. But clearly, it is not true of every set of factors that nomically determines *E* that it contains all and only the causes of *E*. (Moreover, there is no reason for thinking that it is possible to formulate non-trivial necessary and sufficient condition for causation in terms of nomic determination. Philosophers who have tried to do so, typically with reductionist ulterior motives, have been in for a disappointment.)

An assumption that is slightly stronger than (D/d) seems plausible as well:

---

connected to the idea of causes as nomic determiners of their effects (Kment 2014: Sct. 10.4.1, and in particular Sct. 10.4.2), and that this idea can also explain the close connection in ordinary causal thinking between causation by omissions and counterfactual dependence (Kment 2014: Ch. 10).

[17] Admittedly, not all philosophers are happy with the idea that omissions can be causes. For example, Beebee (2004) denies that any omissions are causes, while others hold that they are causes only in a secondary sense, or that they are not causes but stand in some other, closely related relation to effects (Dowe 2000, 2001; Armstrong 2001). Others think that they can be causes in one sense but not in another (Hall 2004). I cannot jump into the fray on this occasion, but see Kment 2014: Scts. 10.4.1–10.4.2, and also Scts. 9.1.2–9.1.3).

[18] Philosophers sometimes distinguish between causes and causally relevant background conditions, or between causes and enablers. However, the term "cause" as used in (D/d) is to be understood in a broader way, as covering all factors that are causally relevant to *E*, including background conditions or enablers. The same is true for the principles (D/d*), (D/i), and (D/i*) below.

(D/d*) Under determinism, the causes of *E* that obtain at *t* nomically determine *E* (where *t* is earlier than the time at which *E* obtains).

The causes of *E* that obtain at *t*—I will call them the '*t*-causes' of *E*—make up a complete temporal cross-section of *E*'s causal history. They nomically determine all later causes of *E* and they screen off any previous causes. (Earlier causes do not act at a temporal distance. They contribute causally to *E* only by causing *t*-causes of *E*.) Hence, if all the causes of *E* together nomically determine *E*, then so do the *t*-causes of *E*. Here is another way of looking at it. Under determinism, the state of the world at *t* contains a set of factors that nomically determines *E*. (D/d*) tells us that, if you remove from the state of the universe at *t* all the factors that are not causally relevant to *E*, then the remaining factors still nomically determine *E*.

It is not of critical importance for my purposes whether the determination idea should be regarded as true in light of our best philosophical and scientific theories.[19] My reconstruction of everyday causal and counterfactual reasoning requires only the premise that the determination idea is commonly used in ordinary explanatory thinking, at least as a working assumption. And that much seems very plausible. Suppose that you made a certain type of cake on two different occasions. The first time it was delicious, the second time it was chalky and unappealing. Then it seems very tempting to say: you must have done something the second time that you did not do the first time and which made the second cake taste chalky. In other words, we can conclude from the fact that the two cakes taste different that the factors that are causally responsible for the taste of the first cake are somewhat different from those responsible for the taste of the second cake. Different effects, therefore different causes. That is the contrapositive of: same causes, same effect. And the latter principle, in turn, is most likely motivated by an application of the determination idea.

3.2 The method of difference and the counterfactual method under determinism

I think that the use of counterfactuals to evaluate causal claims is an extension of another method of causal reasoning, which I will consider first: John Stuart Mill's method of difference (Mill 1956, bk. III, ch. VIII, sct. 2). Let me start with an admittedly highly simplified and idealized description of this procedure.

Scenario 1

t_A	*A B C D*
t_A+1	*E*

[19] However, in Kment 2014: Sct. 10.4 (in particular Sct. 10.4.2) I argue that some of the criticisms that have been leveled at the determination idea are misguided, and that some of them rely on controversial views (that I reject) about what the relata of causation are, for example on the view that they are events (or entities similar to events) rather than facts.

Scenario 2

$$\begin{array}{ll} t & \bar{A}\ B\ C\ D \\ t{+}1 & \bar{E} \end{array}$$

You observe a scenario (Scenario 1) in which the causal factors A, B, C, and D are present at time t_A. A little later, E obtains. You want to know what caused E. Now suppose that you also observe Scenario 2. In Scenario 2, B, C, and D obtain but A does not, and E does not obtain a moment later. You infer from these observations that A is a cause of E in Scenario 1.

In order for this line of reasoning to be justified, you need to assume that the initial states of the two scenarios match each other with respect to all the factors that are causally relevant to whether E obtains at the later time, with the possible exception of A. There must not be any other causally relevant differences between the initial states of the two scenarios. (If there is another such difference, then *that* difference might be what is responsible for the fact that E occurs at the later time in Scenario 1 but not in Scenario 2. Then you cannot blame A for the E's occurrence in Scenario 1.) In other words, the causal factors with respect to which the initial states of the two scenarios match each other—B, C, and D—include all factors obtaining at t_A in Scenario 1 that are causally relevant to E, with the possible exception of A. Equivalently:

(3) A, B, C, and D include all the factors that are t_A-causes of E in Scenario 1.

I propose that we reconstruct the method of difference as follows. By assumption (3), the set of E's t_A-causes must be a subset of $\{A, B, C, D\}$. However, you are not sure whether *all* members of this set are causes of E or only some of them. In particular, you do not know whether A is a cause of E. Now you observe Scenario 2. In this scenario, B, C and D occur but E does not occur a moment later. That shows that

B, C and D do not nomically determine E.

However, according to (D/d*), the facts that are t_A-causes of E in Scenario 1 taken together *must* nomically determine E. Hence:

(4) B, C and D do not include all the t_A-causes of E in Scenario 1.

From (3) and (4) you can infer that A is a cause of E in Scenario 1.

For an illustration of this form of reasoning, consider once more the example of the previous section in which you tried to bake the same cake on two occasions. The first time it tasted good but the second time it did not. Given the difference in taste, you conclude that you must have prepared the two cakes in somewhat different ways. That is an application of the determination idea: you infer a difference in causes from a difference in effects. You look more closely and discover that you used somewhat different ingredients on the two occasions. The first time you used A, B and C, while the second time you used only B and C. That was the only difference between the two cases. You conclude that your use of ingredient A on the first occasion must have been a cause of the pleasant taste. More sophisticated versions of this procedure are applied in scientific experiments. (In these cases, Scenario 1 is the "experimental condition," Scenario 2 is the "control condition," and *B*, *C*, and *D* are the background factors that the experimenters are controlling for.) However, my discussion will focus on everyday uses of the method.

The method of difference has two characteristic limitations. *Firstly*, it requires you to either find or create a control scenario that matches Scenario 1 in all those t_A-factors that are causally relevant to *E* in Scenario 1. However, you might not be lucky enough to find such a scenario and it may be beyond your powers to create one. *Secondly*, you may not even know what scenario to look for or be able to tell whether you have found what you need. For, you may not know very much about which factors at t_A were causally relevant to *E* in Scenario 1 and therefore may not know in what respects the control scenario needs to match Scenario 1.

If my reconstruction of the method of difference is on the right track, however, then it is easy to solve these two problems. Start with the first problem. Suppose that you can narrow down the range of factors that might be t_A-causes of *E* in Scenario 1 to a fairly small set *X*. However, you are unable to find or create an actualized control scenario that matches Scenario 1 with respect to all the factors in $X–\{A\}$. On my account, this is no serious obstacle. For, I claim that you need a control scenario only in order to show that $X–\{A\}$ does not nomically determine *E*. An *actualized* control scenario is not needed to show this. You can instead consider a *possible* scenario where the factors in $X–\{A\}$ obtain and which conforms to the actual laws. If you can show that *E* does not obtain in that possible scenario, then you can conclude that $X–\{A\}$ does not nomically determine *E*. The rest of the argument proceeds in the way described before. This shift from looking for an actualized control scenario to merely looking for a possible one allows us to solve the second problem as well. Suppose you know very little about what caused *E* in Scenario 1, and are therefore unable to narrow down the range of factors that might (for all you know) have been t_A-causes of *E* to a small set. Then you may have little hope of finding an actualized control scenario that matches Scenario 1 in all t_A-causes of *E* other than *A*. However, this problem disappears once we recognize that a merely possible scenario can serve as control scenario. We can simply use as our control scenario a possible world that matches actuality at t_A in *all* factors other than *A* and that conforms to the actual laws after t_A. Suppose that we can show that *E* fails to obtain at worlds like that, i.e. that the following is true:

(5) *E* fails to obtain at those possible worlds where (i) *A* fails to obtain, (ii) the state of the universe at t_A is otherwise just like in actuality, and (iii) events conform to the actual laws of nature after t_A.

We can infer from (5) that the factors other than A that actually obtain at t_A do not nomically determine E. Given (D/d*), it follows that these factors do not include all the actual t_A-causes of E and that A must therefore be one of E's causes.[20]

(5) is what we express by saying that E would not have obtained if A had not obtained. Counterfactuals allow us to state this intermediate conclusion of the reasoning process concisely, and that is likely one of the purposes for which counterfactuals exist. I will call the method of causal reasoning described in this section the "counterfactual method" of supporting causal claims.

3.3 Comparison with John Mackie's account

The account sketched in section 3.2 is similar in some respects to John Mackie's view. Let me briefly compare the two proposals.

In *The Cement of the Universe*, his seminal study of causation, Mackie (1974: chs. 1–3) aims to answer two questions: "What do causal claims mean?", and "What is causation 'as it exists in the objects'?" (Mackie 1974: 60). On his account of the meaning of causal claims, the content of "A is a cause of E" includes certain counterfactual conditionals, such as the claim that E would not have occurred if A had not occurred. Among the "grounds" (ibid.) of these counterfactuals is a certain fact about the actual world, namely the fact that A is a member of a set of actual conditions that are minimally sufficient for E (Mackie 1974: ch. 3).[21] Mackie holds that this fact is part of what constitutes causation as it exists in the objects. He goes on to discuss how the method of difference can be used to show that A is part of a minimal sufficient condition for E (Mackie 1965, 1974: ch. 3). We need to start from some assumptions about Scenarios 1 and 2, including the premise that E has a cause (and that there is therefore a minimal sufficient conditions for E) in Scenario 1, and that the two scenarios are alike in all relevant factors except A. Given E's absence in

[20] Again, this is a little simplified. Let S be the set of facts about t_A other than A. As mentioned in footnote 8, the closest worlds where A fails to obtain do not match actuality with respect to *all* fact in S. Other things being equal, they match actuality as closely in S-facts as is compatible with A's failure to obtain, but there might be a small range of S-facts that fail to obtain at these worlds. Consequently, the inference from the premise that

> E fails to obtain at the closest worlds where A fails to obtain

to the conclusion that

> the facts in S do not nomically determine E (and therefore do not include all the t_A-causes of E)

is defeasible. The premise might be true and the conclusion false if S includes facts that nomically determine E but some of these facts fail to obtain at the closest worlds where A fails to obtain. A fuller version of my account therefore predicts that the inference from E's counterfactual dependence on A to the claim that A is a cause of E is defeasible, or in other worlds, that counterfactual dependence is not quite a sufficient condition for causation. I think that this prediction is borne out (see footnote 7). A fully developed version of the view propounded in this paper can explain why the inference from counterfactual dependence to causation fails in just those cases where it does (see Kment 2014: Sct. 12.1).

[21] I am simplifying by ignoring the fact that Mackie is relativizing such causal claims to a "causal field," which is essentially a set of background factors.

Scenario 2, there can be no sufficient conditions for *E* in Scenario 2. It follows that in Scenario 1, any sufficient conditions for *E* includes *A*, and that *A* is therefore part of a minimal sufficient condition for *E*. That in turn supports the counterfactual component of what is asserted by the claim that *A* is a cause of *E* in scenarios similar to Scenario 1. In this way, the observation of Scenarios 1 and 2 can provide support for this causal claim.

Mackie's account of how the method of difference works rests on an elaborate theory that aims to provide necessary and sufficient conditions for causation. By contrast, my own explanation of the method merely assumes that there is a certain *necessary* condition for causation: certain factors are the *t*-causes of E only if they nomically determine E. Moreover—and this is crucial for the topic of this paper—Mackie's account of the connection between counterfactuals and causal claims is completely different from mine. He does not use his account of the method of difference to explain the connection between counterfactuals and causation. Instead, he thinks that the connection simply consists in the fact that certain counterfactuals are part of the content of a causal claim. In my view, by contrast, counterfactuals are not part of what is said by a causal claim. They merely express the intermediate conclusions of a common way of supporting causal claims that is justified in essentially the same way as the method of difference.

3.4 Limitations of the counterfactual method

We saw in section 2 that there are well-known cases (such as those of over-determination and preemption) in which effects do not counterfactually depend on their causes. As mentioned above, that creates a challenge for any attempt to formulate necessary and sufficient conditions for causation in counterfactual terms and therefore for the counterfactual analysis of causation. But it presents no serious difficulty for the view outlined in this paper. What it shows is simply that the counterfactual method (or at least the version of it discussed in this paper[22]) is more useful for supporting causal claims than for refuting them. If we can show that *E* counterfactually depends on *A*, then that supports the claim that *A* is a cause of *E*. But if *E* is counterfactually independent of *A*, then that does not provide similarly strong evidence for the claim that *A* is *not* a cause of *E*, since the case at hand may involve over-determination or preemption.

That is just what we would expect on my account. In fact, as some authors have noted (Mackie 1965, in particular sct. 5, Strevens 2007), the datum can be explained by the earlier

[22] I have only described the simplest way of using counterfactuals to evaluate causal claims. More sophisticated methods may proceed by determining not only whether *E* counterfactually depends on *A*, but also whether *A* and *E* are linked by certain more complex patterns of counterfactual dependencies. (See Pearl 2009, in particular chs. 7–8, Woodward 2003, and the papers cited in footnote 7 as propounding sophisticated forms of the counterfactual analysis.) While it is open to doubt whether any complex pattern of counterfactual dependencies is necessary and sufficient for causation, some such patterns might come much closer to being necessary and sufficient than simple counterfactual dependence. The fact that the relevant patterns fail to hold between *E* and *A* might then lend (strong but defeasible) support to the claim that *A* is not a cause of *E*, even if it does not entail the latter claim.

observation that the determination idea states merely a necessary but not a sufficient condition for a set to contain all the t_A-causes of E. If a set does not nomically determine E, then it does not contain all of E's t_A-causes. But if the set does nomically determine E, nothing interesting follows. In particular, it does not follow that the set contains all t_A-causes of E. Apply this to the counterfactual method. Let S be the set of all factors that obtain at t_A other than A. If E counterfactually depends on A, then S does not nomically determine E. Given the determination idea, it follows that S does not contain all the t_A-causes of E, so that A must be a cause of E. But if E is counterfactually *independent* of A, then the most we can conclude is that S does nomically determine E. However, that does not entail that S contains all the t_A-causes of E or that A is not a cause of E.

We would expect, therefore, that there is causation without counterfactual dependence whenever there are factors at t_A that don't include all of E's t_A-causes but that nevertheless nomically determine E. That is the case in over-determination and preemption scenarios. Consider first an *over-determination* case. Fred's rock and Susie's rock simultaneous hit the window, each causing sufficient damage to break it. Consider the set of all matters of particular fact that obtain at the time t of Susie's throw, except for her throw itself. This set does not contain all t-causes of the window shattering, since it does not contain Susie's throw. But the set nomically determines the window shattering. For, it contains Fred's throw, as well as background facts that nomically determine that his rock will hit the window with sufficient force to break it. Since all these factors obtain at the closest worlds where Susie does not throw her rock, the window breaks at these worlds. The shattering does not counterfactually depend on Susie's throw.

Similarly in cases of preemption. Suppose that Susie throws her rock first and shatters the window. Fred, who intended to break the window, sees that the job has already been done and walks away. Consider the set of all matters of particular fact obtaining at the time t of Susie's throw, except for her throw itself. This set does not contain all the causes of the window's shattering, since it does not contain Susie's throw. But it does nomically determine the window's shattering. For the set contains Fred's intention to shatter the window, as well as background facts that nomically guarantee that nothing will prevent him from carrying out his intention except for something else's shattering the window first. Given that these factors obtain at the closest worlds where Susie does not throw her rock, the window breaks at these worlds. The window shattering does not counterfactually depend on Susie's throw.

4. The counterfactual method under indeterminism

Under pervasive indeterminism, we can almost never establish that A is a cause of E by showing that E counterfactually depends on A, since effects rarely or never counterfactually depend on their causes. Perhaps we can show that E's *chance* counterfactually depends on A, but the inference from this observation to the claim that A is a cause of E is problematic, as is shown by Schaffer's example of section 2. (The chance of the prince's turning into a frog depends counterfactually on Morgana's spell, despite the fact that her spell is not a cause.) However, from the fact that *E*'s chance counterfactually depends on *A*, we *can* infer that *A* is a cause of the fact that *E* had a certain chance. That is to say, if *E* had chance *p* at *t* ("$ch_t(E) = p$," for short) and we can show that

> *E* would not have had chance *p* at *t* if *A* had not obtained,

then we can conclude that

> *A* is a cause of the fact that $ch_t(E) = p$.[23]

This version of the counterfactual method, just like the deterministic variant, rests on a certain version of the determination idea. This version is restricted to the causes of one special kind of fact, namely facts about chances. I will call it (D/i), for "determination idea/indeterministic version." It says that

(D/i) The causes of the fact that $ch_t(E) = p$ jointly nomically determine that $ch_t(E) = p$.

This principle seems very plausible. If *E* has a certain chance at *t*, then there must be some matters of particular fact that causally determine that *E* has that chance at *t*. A slightly stronger version of this principle seems plausible as well:

(D/i*) Those causes of the fact that $ch_t(E) = p$ that obtain after t^* nomically determine that $ch_t(E) = p$ (for any time t^* before *t*).

The causes of the fact that $ch_t(E) = p$ that obtain after t^* screen off earlier causes. (Earlier causes of the fact that $ch_t(E) = p$ do not act at a temporal distance. They influence *E*'s chance at *t* only by way of influencing what happens between t^* and *t*.) It follows that, if (D/i) is true, then (D/i*) is true as well.

Now suppose that we can show the following:

(6) $ch_t(E) \neq p$ at every possible world *w* that meets the following conditions:

(a) *A* fails to obtain at *w*,

(b) *w* is otherwise like actuality at t_A,

(c) *w* matches actuality after t_A in all matters of particular fact that are not actually caused by *A*, and

(d) *w* conforms to the actual laws after t_A.

From (6) we can infer the following:

---

[23] This indeterministic version of the counterfactual method of evaluating causal claims is subject to the same limitations as the deterministic version: in cases of over-determination and preemption, the fact that $ch_t(E) = p$ may fail to depend counterfactually on *A* despite the fact that *A* is a cause of the fact that $ch_t(E) = p$. This limitation can be explained in the way discussed in the previous section.

(7) Those post-t_A factors that are not caused by A do not nomically determine that $\mathrm{ch}_t(E) = p$.

(D/i*) entails this:

(8) The causes of the fact that $\mathrm{ch}_t(E) = p$ that obtain after t_A nomically determine that $\mathrm{ch}_t(E) = p$.

From (7) and (8) we can infer the following:

(9) The causes of the fact that $\mathrm{ch}_t(E) = p$ include some factors that were caused by A.

Finally, given the assumption that causation is transitive (if A is a cause of B and B is a cause of C, then A is a cause of C),[24] (9) entails the conclusion:

A is a cause of the fact that $\mathrm{ch}_t(E) = p$.

Again, (6) is what we express by saying that E would not have had chance p at t if A had not obtained. Counterfactuals mediate the inference to the causal conclusion and that is one of the ways in which counterfactuals are of use to us.

5. Conclusion

Counterfactuals play a central role in causal reasoning. Counterfactual analyses of causation give one explanation of this fact, my account presents another. Let me conclude by briefly comparing the two approaches in light of the results of the earlier sections.

Over-determination and preemption cases show that counterfactual dependence between distinct matters of particular fact is not a necessary condition for causation. That presents a serious problem for the counterfactual analysis of causation, given that the viability of that account depends on the possibility of finding necessary and sufficient conditions for causation cast in counterfactual terms. The same examples present no difficulty for the view outlined in this paper, according to which counterfactual dependence

[24] It is somewhat controversial whether causation is transitive. For discussion of this question, see McDermott (1995), Paul (2000), Hitchcock (2001), Hall (2004, 2007), Lewis (2004a), Paul and Hall (2013: ch. 5), and Kment (2014: Sct. 12.4). If you believe that causation fails to be transitive, you can easily adjust the account I gave of counterfactuals and the counterfactual method to this background belief of yours. You just need to replace all talk about causation with talk about the ancestral relation of causation. For illustration, consider how the counterfactual method under indeterminism would need to be revised. (The deterministic version of the method could be revised in an analogous way.) To begin with, (2)(c) and (6)(c) need to be replaced with the claim that w matches actuality after t_A in all matters of particular fact to which A does not actually stand in the ancestral relation of causation. The revised counterfactual method is a procedure for showing that

(10) A stands in the ancestral relation of causation to the fact that $\mathrm{ch}_t(E) = p$.

The method starts by showing that the reformulated version of (6) is true. From that result, one can infer that those post-t_A factors to which A does not stand in the ancestral relation of causation do not nomically determine that $\mathrm{ch}_t(E) = p$. Given (D/i*), it follows that the post-t_A causes of E include some factors to which A stands in the ancestral relation of causation. That in turn entails (10). Counterfactuals can be used to express the reformulated version of (6) in a concise manner and are therefore useful to us in applying the revised version of the counterfactual method.

merely provides evidence for causal connections but does not constitute them. All we need to conclude from the data is that the counterfactual method is useful mostly for establishing causal claims, not for refuting them. What is more, the account predicts this limitation of the counterfactual method, and explains it by appealing to the fact that the determination idea provides merely a necessary but not a sufficient condition for certain factors to include all the causes of E. Hence, far from presenting a difficulty for the theory, the examples of over-determination and preemption confirm the account.

The need for causal notions in the theory of counterfactuals threatens the counterfactual analysis of causation with circularity. But it does not constitute a problem for the view that counterfactual reasoning is useful for supporting causal claims. All we need to conclude is that counterfactual thinking cannot in general create causal knowledge from scratch. Given that causal notions figure in the truth-conditions of counterfactuals, we cannot generally acquire causal knowledge by counterfactual reasoning unless we have some causal knowledge to begin with. But there is no circularity and no regress. For, the causal knowledge required for our counterfactual reasoning is different from that which we acquire as a result of it (see Kment 2014: Scts. 10.6.2, 11.5 for more detail). We use one item of causal knowledge to gain another.[25] In that way, counterfactual reasoning extends our stock of causal knowledge. And that is what makes it useful.

Finally, by portraying the use of counterfactuals in establishing causal claims as an extension of Mill's method, my view provides a unified account of the workings of the counterfactual method and of the method of difference (including the method of controlled experiments). This theoretical unification is a further virtue of the account.[26]

References

Adams, E. (1975). *The Logic of Conditionals*. Dordrecht: Reidel.

Ahmed, A. (2014). *Evidence, Decision and Causality*. Cambridge: Cambridge University Press.

Armstrong, D. M. (2001). "Going Through the Open Door Again: Counterfactual versus Singularist Theories of Causation," in Gerhard Preyer (ed.), *Reality and Humean Supervenience: Essays on the Philosophy of David Lewis*. New York: Rowman and Littlefield, 163–176.

---

[25] More specifically, in establishing that the fact that $ch_t(E) = p$ counterfactually depends on A, we need to draw on knowledge about which facts about specific post-t_A events were actually caused by A (only those that were not actually caused by A can be held fixed). Once we have established the counterfactual dependence, we can infer a new causal claim: A is a cause of the fact that $ch_t(E) = p$.

[26] This paper is based on a talk I gave at the Linguistic Perspectives on Causation workshop at the Hebrew University. I am grateful to the organizers of this workshop, Nora Boneh and Elitzur Bar-Asher Siegal, to the other workshop participants, and to two referees for this volume. I am also indebted to Oxford University Press for permission to use passages from my book *Modality and Explanatory Reasoning* (Oxford University Press, 2014), and to Wiley for permission to use passages from my paper "Causation: Determination and Difference-Making," *Noûs* 44: 80–111 (© 2010 Wiley Periodicals, Inc.).

Beebee, H. (2004). “Causing and Nothingness,” in J. Collins, N. Hall, and L. Paul (eds), *Causation and Counterfactuals*. Cambridge, MA: MIT Press, 291–308.
Bennett, J. (1974). “Review of Counterfactuals,” *The Canadian Journal of Philosophy* 4: 381–402.
Bennett, J. (1984). “Counterfactuals and Temporal Direction.” *Philosophical Review* 93: 57–91.
Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.
Dowe, P. (2000). *Physical Causation*. New York: Cambridge University Press.
Dowe, P. (2001). “A Counterfactual Theory of Prevention and ‘Causation’ by Omission,” *Australasian Journal of Philosophy* 79: 216–226.
Edgington, D. (2003). “Counterfactuals and the Benefit of Hindsight,” in P. Dowe and P. Noordhof (eds), *Causation and Counterfactuals*. London: Routledge.
Edgington, D. (2011). “Causation First: Why Causation is Prior to Counterfactuals,” in Hoerl et al. 2011: 230–241.
Fine, K. (1975). “Review of Counterfactuals,” *Mind* 84: 451–458.
Gibbard, A. and W. Harper. (1981). “Counterfactuals and Two Kinds of Expected Utility.” In W. Harper, R. Stalnaker, and G. Pearce (eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: Reidel, 125–162.
Hall, N. (2004). “Two Concepts of Causation,” in J. Collins, N. Hall, and L. Paul (eds), *Causation and Counterfactuals*. Cambridge, MA: Bradford Books, MIT Press, 225–277.
Hall, N. (2007). “Structural Equations and Causation,” *Philosophical Studies 132*: *109–136*.
Halpern, J. Y. and J. Pearl (2005). “Causes and Explanations: A Structural-Model Approach. Part 1: Causes,” *British Journal for the Philosophy of Science* 56: 843–887.
Hiddleston, E. (2005). “A Causal Theory of Counterfactuals,” *Noûs* 39: 632–657.
Hitchcock, C. (2001). “The Intransitivity of Causation Revealed in Equations and Graphs,” *Journal of Philosophy* 98: 273–299.
Hoerl, C., T. McCormack, and S. Beck, eds (2011) *Understanding Counterfactuals, Understanding Causation: Issues in Philosophy and Psychology*. Oxford: Oxford University Press.
Hume, D. (1995), *An Inquiry Concerning Human Understanding*. Upper Saddle River: Prentice Hall.
Jackson, F. (1977). “A Causal Theory of Counterfactuals,” *Australasian Journal of Philosophy* 55: 3–21.
Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
Kment, B. (2006), “Counterfactuals and Explanation,” *Mind* 115: 261–310.
Kment, B. (2010). “Causation: Determination and Difference-Making,” *Noûs* 44: 80–111.

Kment, B. (2014). *Modality and Explanatory Reasoning*. Oxford: Oxford University Press.
Kment. B. (2015). "Modality, Metaphysics, and Method." In C. Daly (ed.), *The Palgrave Handbook of Philosophical Methods*, New York: Palgrave Macmillan, 179–207.
Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
Lewis, D. (1981). "Causal Decision Theory." *Australasian Journal of Philosophy* 59: 5–30.
Lewis, D. (1986a). "Causation," in *Philosophical Papers*, vol. 2. New York and Oxford: Oxford University Press, 159–213.
Lewis, D. (1986b). "Counterfactual Dependence and Time's Arrow," in *Philosophical Papers*, vol. 2. New York and Oxford: Oxford University Press, pp. 32–66.
Lewis, D. (2004a). "Causation as Influence," in J. Collins, N. Hall, and L. Paul (eds), *Causation and Counterfactuals*. Cambridge, MA: Bradford Books, MIT Press, 75–107.
Lewis, D. (2004b). "Void and Object," in J. Collins, N. Hall, and L. Paul (eds), *Causation and Counterfactuals*. Cambridge, MA: Bradford Books, MIT Press, 277–290.
Mackie, J. L. 1965, "Causes and Conditions," *American Philosophical Quarterly* 2, pp. 245-64.
Mackie, J. L. 1974, *The Cement of the Universe*, Oxford: Oxford University Press.
Maudlin, T. (2004). "Causation, Counterfactuals, and the Third Factor." In J. Collins, E. J. Hall and L. A. Paul (eds.), *Causation* and *Counterfactuals*. Cambridge, MA: MIT Press.
Mårtensson, J. (1999). *Subjunctive Conditionals and Time*. Acta Universitatis Gothoburgensis. Gothenburg: University of Gothenburg.
McDermott, M. (1995). "Redundant Causation," *British Journal for the Philosophy of Science* 46: 523–544.
Menzies, P. (1989). "Probabilistic Causation and Causal Processes: A Critique of Lewis," *Philosophy of Science* 56: 642–663.
Mill, J. S. (1956). *A System of Logic, Ratiocinative and Inductive*. London and New York: Longmans, Green.
Paul, L. A. (2004). "Aspect Causation," in J. Collins, N. Hall, and L. Paul (eds), C*ausation and Counterfactuals*. Cambridge, MA: Bradford Books, MIT Press, 205–224.
Paul, L. A. and N. Hall (2013), *Causation: A User's Guide*. Oxford: Oxford University Press.
Pearl, J. 2009, *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.
Quine, W. V. O. (1950). *Methods of Logic*. New York: Holt, Rinehart, and Winston.
Ramachandran, M. (1997). "A Counterfactual Analysis of Causation," *Mind* 151: 263–277.
Roese, N. J., and J. M. Olson, eds (2014), *What Might Have Been: The Social Psychology of Counterfactual Thinking*, New York: Psychology Press.

Schaffer, J. (2000). “Overlappings: Probability-Raising Without Causation,” *Australasian Journal of Philosophy* 78: 40–46.

Schaffer, J. (2004). “Counterfactuals, Causal Independence and Conceptual Circularity.” *Analysis* 64: 299–309.

Schulz, K. (2011), “If you’d wiggled A, then B would've changed.” *Synthese* 179: 239–51.

Slote, M. (1978). “Time in Counterfactuals,” *Philosophical Review* 87: 3–27.

Stalnaker, R. (1981). “Letter to David Lewis.” In W. Harper, R. Stalnaker, and G. Pearce (eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: Reidel: 151–152.

Strevens, M. (2007), “Mackie Remixed,” in Campbell, J. K., M. O’Rourke, and H. S. Silverstein, eds., *Causation and Explanation*, Cambridge, MA: MIT Press.

Tichý, P. (1976). “A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals,” *Philosophical Studies* 29: 271–273.

Veltman, F. (2005). Making counterfactual assumption. *Journal of Semantics,* 22: 159–180.

Wasserman, R. (2006). “The Future Similarity Objection Revisited,” *Synthese* 150: 57–67.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Yablo, S. (2004). “Advertisement for a Sketch of an Outline of a Proto-Theory of Causation,” in J. Collins, N. Hall, and L. Paul (eds), *Causation and Counterfactuals*. Cambridge, MA: Bradford Books, MIT Press, 119–138.