Prevention, Preemption, and the Principle of Sufficient Reason

Christopher Hitchcock
California Institute of Technology

1. Counterfactual Theories of Causation

One event, *e*, counterfactually depends upon another event, *c*, just in case *e* would not have occurred had *c* not occurred. Beginning with the seminal paper of David Lewis in 1973, there has been a lively philosophical tradition of trying to analyze token causation in terms of counterfactual dependence. The simplest possible counterfactual theory of token causation—henceforth the *simple theory*—would identify token causation with counterfactual dependence: *c* is a token cause of *e* just in case *e* counterfactually depends upon *c*. This simple account is threatened by counterexamples on both sides. Some authors, but by no means all, take cases of prevention and omission[1] to show that there can be counterfactual dependence without token causation.[2] Cases of *preemption* have been widely taken to show that there can be token causation without

For comments and discussion, I would like to thank Branden Fitelson, Clark Glymour, Ned Hall, Joseph Halpern, Dan Hausman, Franz Huber, Laurie Paul, Jonathan Schaffer, James Woodward, Stephen Yablo, Jiji Zhang, audience members at the University of California at Berkeley, and two anonymous referees for the *Philosophical Review*.

1. Although it is common to talk of 'causation by omission', I omit the words 'causation by' so as not to beg the question as to whether cases of omission are cases of genuine causation. I will sometimes put words such as 'causation' and 'causal' in scare quotes when I intend to include prevention and omission within their scope.

2. Lewis (2000, sec. 10) and Schaffer (2000a), for example, have argued that these are cases of genuine causation; others, Beebee (2004), for example, argue that they are not; whereas Dowe (2000, chap. 6) and Hall (2004), for example, relegate these cases to a kind of secondary causal status.

Philosophical Review, Vol. 116, No. 4, 2007
DOI 10.1215/00318108-2007-012

counterfactual dependence; many authors (but not Lewis himself) also consider cases of *overdetermination* to be counterexamples to the necessity of counterfactual dependence for token causation. There have been many attempts to deal with the problems of preemption and overdetermination,[3] none entirely satisfactory. We will examine the shortcomings of some of these theories in sections 11–13.

In this essay, I will propose a condition that specifies when counterfactual dependence is both necessary and sufficient for token causation. I will develop my account using structural equations and directed graphs to model various causal systems. These tools will be briefly explained in section 3. Within this framework, I will formulate a principle that I will call the *Principle of Sufficient Reason*. This principle captures an important feature of our causal reasoning. This principle will, in turn, be used to define the notion of a *self-contained network*: a network is self-contained if and only if it satisfies the Principle of Sufficient Reason. My central proposal is this: when a network is self-contained, then our judgments of token causation are strongly aligned with counterfactual dependence. That is, when e counterfactually depends upon c in a self-contained network, then we are strongly inclined to judge that c is a token cause of e; and when e does not counterfactually depend upon c in a self-contained network, then we are strongly inclined to judge that c is not a token cause of e. It is only when we have a causal network that is not self-contained that we are forced to abandon the simple theory. The result is a rather disjunctive theory of token causation. When a causal network is self-contained, counterfactual dependence is both necessary and sufficient for causation. When a causal network is not self-contained, counterfactual dependence indicates a kind of secondary dependence that I will call *parasitic dependence*. I will remain agnostic on whether or not parasitic dependence is genuine causation: cases of parasitic dependence are, at any rate, interestingly different from paradigmatic examples of token causation. Prevention and omission fall into this category. Finally, when we have neither self-containment nor counterfactual dependence, we must appeal to one of the more sophisticated counterfactual theories alluded to above to assess whether or not we have token causation. This

3. An incomplete list would include: Barker 2004; Björnsson 2007; Colton (n.d.); Ganeri, Noordhof, and Ramachandran 1996, 1998; Glymour and Wimberly 2007; Halpern and Pearl 2001, 2005; Hiddleston 2005; Hitchcock 2001, 2004b; Lewis 1973, 1986a, 2000; McDermott 1995; Menzies 2004a, 2004b; Noordhof 1999, 2004; Pearl 2000, chap. 10; Ramachandran 1997, 1998, 2004a, 2004b; Woodward 2003, chap. 2; and Yablo 2002, 2004.

framework is then applied to a number of different kinds of problem case: prevention, omission, affecting, background conditions, early and late preemption, transitivity, and symmetric overdetermination.

2. Preliminaries

In this section, I will address a number of preliminary issues, do some stage setting, and clarify some of the terminology that is to be used in this essay.

By 'token causation' I mean the sort of causal relationship that is reported in claims such as "Assassin's poisoning the coffee caused Victim to die." This is the relationship that is the target of analysis in Lewis 1973 and Lewis 2000 and in most work in the counterfactual tradition. I take the term 'token causation' to be fairly standard; 'singular causation' and 'actual causation' are also fairly common; just plain 'causation' is doubtless the most common term of all. By using 'token causation' rather than 'causation' simpliciter, I deliberately suggest that the relation being studied here does not exhaust the topic of causation. By contrast, my usage of 'token causation' rather than 'singular causation' or 'actual causation' carries no significance. In particular, I do not take it to be a unique or distinguishing feature of token causation that it relates tokens.

For convenience, I will talk of token causation as a relation whose relata are events. This might be wrong. Mellor (1995, 2004), for example, has argued that causation at least sometimes relates *facts* and that it is not a genuine relation. If this is correct, then the term 'token causation' might be inappropriate since it does not relate *tokens* at all (given that facts are abstracta rather than concrete particulars).[4] Nothing in my account will depend upon the underlying metaphysics, and I encourage those who reject the view that token causation is a relation among events to paraphrase as they see fit. In referring to causes and effects, I will freely switch back and forth between perfect and imperfect nominals. Thus I will take "Assassin's poisoning the coffee caused Victim to die" to be synonymous with "Assassin's poisoning of the coffee caused Victim's death," although to some ears the two might suggest different underlying ontologies.

My proposal will be pragmatic in orientation, along at least two dimensions. First, instead of formulating my account in terms of the familiar distinction between positive events and omissions or absences,

4. Thanks to Dan Hausman for pointing this out.

I will instead introduce a new distinction between *deviant* and *default* outcomes of various processes. This distinction will be discussed in greater detail in section 5. Although the deviant/default distinction frequently tracks the event/absence distinction, it does not do so universally; we will examine an important exception in section 9.

Second, although I will offer an analysis of token causation, my suspicion is that there is no coherent concept that is tracked by our intuitive judgments of token causation; rather, there is a set of disparate factors that tug our judgments in different directions. If this is so, then the most I can hope to achieve is to identify one type of consideration that exerts a strong pull on our judgments. Moreover, our intuitive judgments often have interesting contours that get lost when we summarize them as judgments that one event does or does not cause another. For example, we noted above that there is some difference of opinion regarding whether cases of prevention and omission are cases of genuine causation.[5] What we should demand of a theory of causation is not so much that it settle this disagreement in one way or the other, but that it identify the respects in which cases of prevention and omission both resemble and differ from paradigmatic cases of causation.

My project will be restricted in scope in at least three ways. First, I will take for granted that we can make sense of certain nonbacktracking counterfactuals, in the sense of Lewis 1979. I will not, however, assume that these can be analyzed without causal remainder. Even if nonbacktracking counterfactual dependence is itself a causal notion, the counterexamples to the simple theory show that it is not the *same* causal notion as token causation; hence an account of the latter in terms of the former need not be trivial or unilluminating.

Second, I deal here only with the deterministic case. The addition of probabilities helps with some of the problems addressed here,[6] but it also creates new problems.[7] A proper treatment of these issues must await another occasion.

Third, there are a number of different counterfactual (and related) theories of token causation on offer,[8] and my account bears systematic connections with many of them. Although I will briefly discuss

5. See note 2.

6. See especially Hiddleston 2005 for some interesting suggestions along these lines.

7. See Hitchcock 2004a for a discussion of the most significant problem.

8. See note 3. Menzies' approach is perhaps the most similar to that developed here.

a few of these approaches, a thorough presentation of these alternatives and comparisons with my present account is beyond the scope of this essay.

3. Causal Models

In this section, I will give an example to illustrate the causal modeling techniques that will be used throughout this essay. I attempt here only a brief introduction to these techniques: for details see Pearl 2000; Halpern and Pearl 2001, 2005; Hitchcock 2001, 2004b. The illustration will be an example of *early preemption*, a type of case that will be discussed in greater detail in section 11. For uniformity, all of the examples to be used in this essay will involve poisonings or attempted poisonings. Some writers have decried the violent nature of the examples used in the causation literature, but as any fan of English manor murder mysteries can tell you, poisoning is the preferred modus operandi of the gentleman (or gentlewoman) murderer.

Here is our first example:

> *Early Preemption.* Assassin poisons Victim's coffee. Victim drinks it and dies. If Assassin hadn't poisoned the coffee, Backup would have, and Victim would have died anyway. Victim would not have died if there had been no poison in the coffee.

This is a case of causal preemption: Assassin's action preempts Backup from poisoning Victim. It is interesting because it is a counterexample to the simple theory: we judge that Assassin's poisoning the coffee causes Victim's death even though Victim would have died if Assassin had not poisoned the coffee.

A causal model is an ordered pair <**V**, **E**>, where **V** is a set of variables and **E** is a set of equations among these variables. A variable can take on different values, where each value represents the occurrence (or nonoccurrence) of some event, or perhaps a *version* of some event (to use the terminology of Lewis 2000). To represent *Early Preemption*, it would be natural to use three variables, A, B, and D, having the following interpretations:

A = 1 if Assassin poisons the coffee, 0 if not.
B = 1 if Backup poisons the coffee, 0 if not.
D = 1 if Victim dies, 0 if not.

The variables in a causal model need not be binary in general; the values of a variable could be used to represent a range of possibilities, or perhaps the value of a continuous quantity such as mass. In almost all of the examples we will discuss, however, it will suffice to use binary variables (the exception occurs in section 10).

The description of the case involves certain counterfactuals: If Assassin hadn't poisoned the coffee, Backup would have, and Victim would have died anyway; if neither Assassin nor Backup had poisoned the coffee, then Victim would not have died; and so on. These counterfactuals are represented using *equations* among the variables. The convention is that the variables appearing on the right-hand side of an equation figure in the antecedents of the corresponding counterfactuals, and those appearing on the left figure in the consequents. Each equation asserts several counterfactuals: one for each assignment of values to the variables that makes the equations true. *Early Preemption* can be represented using the following equations:

$$\begin{aligned} EP \quad & A = 1 \\ & B = \sim A \\ & D = A \vee B \end{aligned}$$

(I will adopt the convention of using a full name, such as *Early Preemption*, to denote a hypothetical scenario and an abbreviated name, such as *EP*, to denote its corresponding causal model. I will also use the name of the model to label the set of equations contained in the model, as I have done here, although technically the set of equations is only one component of the model.) Note that I am using symbols familiar from symbolic logic to express mathematical functions of binary variables: $\sim A \equiv 1—A$, $A \vee B \equiv \max\{A, B\}$, and so on. The first equation expresses a "factual" rather than a counterfactual: it asserts that Assassin did in fact poison Victim's coffee. A variable such as A, whose value is given rather than determined by the other variables in the model, is said to be an *exogenous* variable. The second equation encodes two counterfactuals: for the value $A = 1$, it says that if Assassin had not poisoned the coffee, Backup would have; for $A = 0$, it says that if Assassin had poisoned the coffee, Backup would not have. The third equation encodes four counterfactuals. It tells us that Victim would have died if: (i) both Assassin and Backup had poisoned the coffee; (ii) Assassin had, but Backup had not, poisoned the coffee; or (iii) Assassin had not, but Backup had, poisoned the coffee. It also tells us that Victim would have survived if (iv) neither Assassin nor Backup had poisoned the coffee.

Each variable appears on the left-hand side of exactly one equation; this equation expresses the rule whereby the value of that variable is determined. We will assume that all sets of equations are *acyclic*: there is no sequence of variables $X, Y, \ldots, Z$, such that X appears on the right-hand side of the equation for Y, Y appears on the right-hand side of the equation for the next variable, . . . , the penultimate variable appears on the right-hand side of the equation for Z, and Z appears on the right-hand side of the equation for X. This assumption may fail if there are causal loops, but we will ignore this possibility.

Not every counterfactual that is true in the scenario is explicitly represented by an equation; rather, the equations form a minimal generating set for all the true counterfactuals. In *Early Preemption*, for example, if Assassin had not poisoned the coffee, Victim still would have died; and if Victim had not died, then Assassin still would have poisoned the coffee (remember that the counterfactuals do not backtrack). To evaluate any counterfactual whose antecedent specifies the value(s) of one or more variables, we *replace* the equation(s) for the relevant variable(s) with one(s) that stipulates the new value(s) of the variable(s). Let us state this formally:

> *CF* Let <**V**, **E**> be a causal model, and let $X_1, X_2, \ldots, Y_1, Y_2, \ldots \in \mathbf{V}$. The *counterfactual* "If X_1 had been x_1, X_2 had been $x_2, \ldots$; then Y_1 would have been y_1, Y_2 would have been $y_2, \ldots$" *is true in* <**V**, **E**> just in case the following condition holds: In the new causal model <**V**, **E'**> formed by replacing the equations for $X_1, X_2, \ldots$ in **E** with the new equations $X_1 = x_1, X_2 = x_2, \ldots, Y_1$ takes the value y_1, Y_2 takes the value y_2, and so on.

To calculate what would have happened if Assassin had not poisoned the coffee, we replace the first equation in *EP* with $A = 0$. We can then compute that B would have been equal to 1 and D would have been equal to 1. To calculate what would have happened if D had been equal to 0, we replace the third equation with $D = 0$: we do *not* substitute 0 for D in the third equation since that would lead to backtracking. Note that this procedure is consistent with the direct counterfactual interpretation of the equations. The definition of counterfactual dependence follows straightforwardly:

CD Let <**V**, **E**> be a causal model, let $X, Y \in \mathbf{V}$, and let the actual values of X and Y in the model be x and y, respectively. *Y counterfactually depends* upon X in <**V**, **E**> just in case there exist values of X and Y, $x' \neq x$, $y' \neq y$ (respectively) such that "if X had been x', then Y would have been y'" is true in <**V**, **E**>.

It is often helpful to represent a causal model using a directed graph with nodes corresponding to the variables. The graph does not add any information not contained in the model, but often serves to render certain features of the model salient. The convention is that an arrow is drawn from one variable to another just in case the former appears on the right-hand side of an equation with the latter on the left. In such a case we say that the former variable is a *parent* of the latter. The graph for *Early Preemption* is depicted in figure 1. The representation is qualitative in the sense that information about the mathematical forms of the equations is left out of the graph.

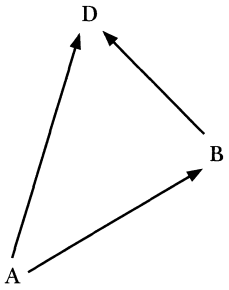

Figure 1.

I finish this section with three further remarks on the nature of causal models. The first is that, in constructing a model, it is important to choose the variables so that different values of the same variable correspond to events (or versions of events) that are incompatible on broadly logical or conceptual grounds; typically, they will represent incompatible states of a system at the same time. For example, it is logically impossible that Assassin poison Victim's coffee and not poison Victim's coffee at the same time, so these two possibilities are appropriate referents for different values of the same variable. A corollary of this principle is that the values of different variables should correspond to events that are *distinct*, in the sense of Lewis 1986b. In *Early Preemption*, Assassin and Backup will not both put poison in Victim's coffee, so there is a sense in which the two poisonings are incompatible. Nonetheless, it would be wrong to model *Early Preemption* using a variable P, such that $P = 1$ corresponds to Assassin's poisoning the drink and $P = 0$ corresponds to Backup's poisoning the drink. These two possible events are distinct from one another: the nature of their incompatibility is not logical or conceptual, but rather

"causal": Assassin's action *prevents* Backup from poisoning the coffee but not vice versa, a feature of the case that will be obscured if we represent the two possible events as values of a single variable. These restrictions on the choice of variables in a model are discussed in greater detail in Hitchcock 2001, 2004b.

The second remark is that although the variables in a causal model *represent* various events that occur or might have occurred, and the equations *represent* patterns of counterfactual dependence among those events, it is often convenient to drop explicit talk of representation. Thus I will say such things as that $A = 1$ occurs, or that A takes the value 1 (rather than that the event represented by $A = 1$ occurs). Most of the time, this contraction will cause no confusion. There is one area, however, where we must take care. I will offer conditions for when $X = x$ is a token cause of $Y = y$ in a causal model $\langle \mathbf{V}, \mathbf{E} \rangle$. Ultimately, however, we are interested in whether one event, c, is a token cause of another, e. My proposal is that c is a token cause of e just in case: (i) $X = x$ is a token cause of $Y = y$ in causal model $\langle \mathbf{V}, \mathbf{E} \rangle$, as defined below; (ii) $X = x$ represents c and $Y = y$ represents e; and (iii) $\langle \mathbf{V}, \mathbf{E} \rangle$ is an *appropriate* causal model of the situation in which c and e occur. What constitutes an appropriate model is a tricky affair, more a matter of art than science. At the very least, an appropriate model must be constructed in accordance with the restrictions described in the previous paragraph, and it must entail only true counterfactuals. More nebulously, it must include enough variables to capture the essential structure of the situation being modeled. What counts as an appropriate model may depend at least in part on pragmatic factors. Given my pragmatic orientation toward the notion of token causation, this level of pragmatic infection does not disturb me. For further discussion, see Hitchcock 2001, 2004b.

The final remark addresses the question of just what is being represented in causal models such as *EP*. Let us start with what is *not* being represented. They do not explicitly represent relations of token causation: one cannot simply read off token causal relations from either the equations or the graph. Nothing in *EP* makes it obvious that $A = 1$ is to count as a cause of $D = 1$, but that $B = 0$ is not. But neither do causal models represent *type*-level causal relations, for example, between poisoning and death. The variables and their values refer to specific events involving specific individuals at specific places and times. Nor do the equations and graphs represent something that is altogether *acausal*: they do not represent relations of mere succession or co-occurrence, for example. I will coin a term, then, and say that each causal model represents the

token causal structure of the situation in question. Although the distinction between token- and type-level causation has received considerable attention, the distinction between token causation and token causal structure has not. (See Hitchcock n.d. for further discussion of this distinction.)

Token causal structure gives us much of what we want from causation. The causal model given above allows us to predict whether Victim will survive, given information about the actions of the other agents. It tells us which kinds of interventions would result in Victim's survival or death. It gives us information about counterfactual scenarios under which the outcome would be different. Token causation, apparently, is not necessary for any of these. Token causation is involved specifically in our post hoc evaluations of responsibility: after the fact, which agent's actions were responsible for the outcome? This notion is important to philosophers since it plays a role in concepts like moral responsibility and singular explanation. Nonetheless, we can afford to let judgments of token causation be infected by pragmatic criteria without giving up on the objectivity of causation generally: objectivity can be retained at the level of token causal structure.

4. The Intuition of Difference

Consider now a case of omission:

> *Omission*: Assassin poisons Victim's coffee. Bodyguard is in possession of an antidote that is capable of neutralizing the poison, but she refrains from administering it to Victim. Victim dies from the poison. Victim would not have died if the antidote had been administered or if the coffee had not been poisoned.

In order to model this scenario, we may choose the following variables:

A = 1 if Assassin poisons Victim's coffee, 0 if not.
B = 1 if Bodyguard administers the antidote, 0 if not.
D = 1 if Victim dies, 0 if not.

The corresponding equations are:

$$\begin{aligned} Om \quad & A = 1 \\ & B = 0 \\ & D = A \,\&\, {\sim}B \end{aligned}$$

The graph is shown in figure 2.

This scenario has two subplots:

Omission-a: Assassin poisons Victim's coffee. Victim dies from the poison. Victim would not have died if the coffee had not been poisoned.

Omission-b: Bodyguard is in possession of an antidote, but she refrains from administering it to Victim. Victim dies. Victim would not have died if the antidote had been administered.

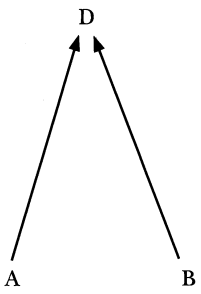

Figure 2.

In terms of counterfactual dependence, Bodyguard's failure to administer the antidote stands in the same relationship to Victim's death as does Assassin's poisoning the coffee. Yet regardless of whether we regard Bodyguard's inaction as a genuine cause of Victim's death, we feel that there is an intuitive difference between the two cases. Dowe (2000, sec. 6.1) appeals to this "intuition of difference" to bolster his position that cases of omission are not cases of genuine causation. What is the difference between the two? Intuitively, *Omission-a* strikes us as a self-contained story. The counterfactual dependence of Victim's death upon Assassin's action is due to the nature of the direct causal connection between the two. This is not to say that the poisoning by itself is a *sufficient condition* for Victim's death. As Mill 1843 taught us, it is only the presence of poisoning together with a host of other conditions—Victim's physiology, the contents of his stomach, prevailing atmospheric conditions, and yes, the absence of a counteracting antidote—that suffice for death. Nonetheless, there is some intuitive sense in which the poisoning provides a satisfactory explanation for Victim's death. By contrast, *Omission-b* strikes us as incomplete. Refraining from administering an antidote is not itself something that induces death. Rather, Victim's death counterfactually depends upon Bodyguard's inaction only because Victim had already been poisoned. The relationship between Bodyguard's inaction and Victim's death (whether it is genuine causation or not) is *parasitic* upon Assassin's action. So *Omission-b* strikes us as incomplete in the absence of any reference to the poison in the coffee.

At this stage, our explanation of the difference between *Omission-a* and *Omission-b* is still very informal. It will be my task to make this idea precise in sections 6 and 8 below.

5. Defaults and Deviants

As I mentioned in section 2, I prefer to speak of default and deviant values of variables rather than of positive events and absences. As the name suggests, the default value of a variable is the one that we would expect in the absence of any information about intervening causes. More specifically, there are certain states of a system that are self-sustaining, that will persist in the absence of any causes other than the presence of the state itself: the default assumption is that a system, once it is in such a state, will persist in such a state. Theory—either scientific or folk—informs us which states are self-sustaining in this way. For example, Newtonian physics tells us that an object's velocity is self-sustaining, whereas its acceleration is not. Thus the default is that the object will maintain the same velocity. The default may depend upon the level of analysis. Consider, for example, a variable whose values represent the state of an individual—alive or dead. It is a plausible principle of folk biology that an individual will remain alive unless something causes her to die, hence it would treat 'alive' as the default value of the variable. But from the perspective of a physiologist, remaining alive requires an amazing effort on the part of complex, delicate systems, as well as interactions with the environment; hence death might be viewed as the default state. Perhaps a case could be made for allowing only genuine laws of nature to determine default values of variables,[9] but if we disallow folk theories, we are not likely to arrive at a theory that accords with folk intuitions. Note also that the default value of a variable may not be an intrinsic feature of the state that is represented. That is, we could have two individuals in the very same state, while one is in a deviant state and the other in a default state. For example, in *Early Preemption* and *Omission*, it is natural to take the default value of the variable D to be 0; the default is that Victim will be alive. But suppose we construct a causal model to represent the story in which Jesus raises Lazarus from the dead. Such a model would include a variable representing Lazarus's state at the end of the scenario (alive or dead), and here it would be natural for the default to be that Lazarus is dead: we expect Lazarus to remain dead in the absence of Jesus' intervention. In other words, the default is not that one *is* alive or dead, but rather that one will *remain* in a state of being alive or dead, depending on how one started the day.

9. Such a view might well make contact with approaches to causation in terms of conservation laws, such as Dowe 2000. See also Maudlin 2004 for an argument that laws should play a role in causal analysis by determining default states.

In addition to this definition of default and deviant values of a variable, I offer several rules of thumb. Temporary actions or events tend to be regarded as deviant outcomes. In the case of human actions, we tend to think of those states requiring voluntary bodily motion as deviants and those compatible with lack of motion as defaults. In addition, we typically feel that deviant outcomes are in need of explanation, whereas default outcomes are not necessarily in need of explanation. Frequently, but not always, my deviant values correspond to positive events, and defaults correspond to absences or omissions.

In most cases, the assignment of default and deviant values is fairly straightforward. In the cases described above, it is natural to assume that Assassin's administering the poison, Backup's administering the poison, Bodyguard's administering the antidote, and Victim's death are all deviant outcomes. When we construct a causal model <**V**, **E**>, it will be necessary to specify the default and deviant values of the variables in **V**, which we will do using the following notation: $\text{Def}(X) = x$, $\text{Dev}(X) = x'$. Nonbinary variables may have multiple deviant or default values, so strictly speaking, the functions Def and Dev take sets as values, not individual values, but I will not distinguish between singleton sets and their members in what follows. I will typically adopt the convention of using the value 0 to represent the default value of a variable, and 1 to represent the deviant value; this convention has been used in all of the examples introduced so far. When this convention is being employed, I will not explicitly specify the deviant and default values.

Finally, there is one difficult kind of case. Suppose that our causal model includes variables that represent the velocity of an object at different times, V_t and $V_{t'}$ with $t' > t$. What is the default value of $V_{t'}$? If Aristotelian physics were true, the default value would be 0: an object's natural state is to be at rest unless something is compelling it to move. In Newtonian physics, however, there is no default velocity; rather, an object will maintain a constant velocity unless acted upon by a net force. Thus the default velocity of the object at time t' is whatever velocity the object had at earlier time t. We will say that a variable such as $V_{t'}$ is an *inertial* variable and express its default as $\text{Def}(V_{t'}) = V_t$. Inertial variables will play an important role in the analysis of late preemption in section 13.

6. The Principle of Sufficient Reason and Self-Contained Networks

As the names suggest, we reason differently about default and deviant outcomes. I will try to capture this difference in a principle that I will

call, with no pretense of historical accuracy, the *Principle of Sufficient Reason*. The best-known formulation of this principle is of course due to Leibniz (1956 [1716], sec. 2):

> nothing happens without a sufficient reason, why it should be so, rather than otherwise.

My proposal is to restrict the scope of the quantifiers 'nothing' and 'a' to *deviant* values of variables. John Stuart Mill (1843, vol. 1, chap. 5, sec. 3) offers a statement that comes close to what I have in mind:

> From nothing, from a mere negation, no consequences can proceed. All effects are connected, by the law of causation, with some set of *positive* conditions.

The idea, in my terminology, is that when a set of variables all take their default value, they cannot by themselves cause another variable to take a deviant value. More formally:

> *PSR* Let <**V**, **E**> be a causal model, and let $X \in \mathbf{V}$. If X has parents in <**V**, **E**>, then, when every parent of X in <**V**, **E**> takes a default value, X takes a default value.

Note that a variable can satisfy *PSR* trivially if it has no parents (that is, when it is exogenous). The contrapositive of *PSR* states that if X takes a deviant value, and X has at least one parent in <**V**, **E**>, then at least one parent of X will take a deviant value. *PSR* is, I believe, an extremely natural principle of causal reasoning. For example, note that something like *PSR* is assumed in the standard conventions for drawing "neuron diagrams": a neuron will fire only if some other neuron "stimulates" it by firing.[10]

I will not, however, impose *PSR* as a general constraint on the variables in a causal model. A causal model is always incomplete in one way or another. Nonetheless, models whose variables all satisfy *PSR* are self-contained in a way that other models are not. If the variable X in causal model <**V**, **E**> does not satisfy *PSR*, then we implicitly understand that there must be some variable, excluded from **V**, whose deviant value(s) explain why X takes on deviant value(s). I will elevate this intuition to the status of a definition:

10. See Hitchcock 2007 for further discussion of neuron diagrams and their connection with structural equation models.

SC A causal model <**V**, **E**> is *self-contained* just in case every variable $X \in \mathbf{V}$ satisfies *PSR*.

Let's look at our examples. The causal model *EP* contains the equation $B = \sim A$. Thus, when A takes its default value of 0, B takes the deviant value 1. Since A is the only parent of B in this causal model, B does not satisfy *PSR*. Hence *EP* is not self-contained. *Om*, by contrast, is self-contained: the variables A and B have no parents, and thus satisfy *PSR* trivially, while D takes the deviant value 1 only when its parent A does.

The concept that we will need is not quite that of a self-contained model, however, but rather that of a self-contained *causal network*. In order to define this concept, we will need two other definitions. We start with the notion of a *directed path*:

DP Let <**V**, **E**> be a causal model, and let $\{X_1, \ldots, X_n\} \subseteq \mathbf{V}$. $\{X_1, \ldots, X_n\}$ constitutes a *directed path* in <**V**, **E**> if the variables $X_1, \ldots, X_n$ can be arranged in a sequence $<X_{i1}, \ldots, X_{in}>$ such that each X_{ij} is a parent of X_{ij+1}, for $j = 1, \ldots, n—1$. Such a directed path will be said to be a *path from* X_{i1} *to* X_{in}.

Although the notion of an undirected path plays a role in some causal modeling techniques, I will not make use of that notion here, and hence will use the simple term 'path' to mean 'directed path'. The notion of a path is most easily understood using causal graphs: a path is a set of variables that are all connected by a series of arrows that meet tip to tail. For example, in causal model *EP* above, $\{A, D\}$ and $\{A, B, D\}$ are both paths from A to D. A causal network is then the system of all paths connecting two variables in a causal model:

CN Let <**V**, **E**> be a causal model, and let $X, Y \in \mathbf{V}$. The *causal network* connecting X to Y in <**V**, **E**> is the set $\mathbf{N} \subseteq \mathbf{V}$ that contains exactly X, Y, and all variables $Z \in \mathbf{V}$ lying on a path from X to Y in <**V**, **E**>.

For example, in *EP*, the causal network connecting A to D is $\{A, B, D\}$—all of the variables in *EP* lie on a path from A to D. In *Om*, the causal network connecting A to D is just $\{A, D\}$, and the causal network connecting B to D is $\{B, D\}$.

We are now ready to define a self-contained network:

SCN Let <**V**, **E**> be causal model, and let X, $Y \in \mathbf{V}$. Let $\mathbf{N} \subseteq \mathbf{V}$ be the causal network connecting X to Y in <**V**, **E**>. Then the causal network **N** is self-contained if and only if for all Z in **N**, if Z has parents in **N**, then Z takes a default value when all of its parents in **N** do (and its parents in $\mathbf{V} \setminus \mathbf{N}$ take their actual values).

That is, **N** is self-contained if every variable in **N** satisfies a restricted version of *PSR*, where only parents that are themselves in **N** are relevant. Intuitively, a network is self-contained when it is never necessary to leave or augment the network to explain why the variables within the network take the values that they do. When a variable (other than the first) in a self-contained network takes a deviant value, this can be explained in terms of the deviant value of one or more of its parents in the network.

Let's illustrate the concept of a self-contained network using our examples. In *EP*, the causal network connecting A to D, $\{A, B, D\}$, is not self-contained. The variable B lies in this network, yet when its lone parent in the network, A, takes the default value 0, B takes the deviant value 1. In *Om*, the causal network connecting A to D, $\{A, D\}$, is self-contained. A has no parent in the network, so it meets the condition trivially. D has one parent in the network, namely A, and D takes the default value 0 when A does. By contrast, the causal network connecting B to D in *Om*, namely $\{B, D\}$, is not self-contained. D contains one parent in the network, B, and when B takes its default value of 0, D takes the deviant value of 1. This difference between the two causal networks in *Om* captures the intuition of difference described in section 4. *Omission-a*, the story in which Assassin causes Victim's death by poisoning the coffee, strikes us as a satisfactory, stand-alone account of why Victim died. The corresponding causal network is self-contained. By contrast, *Omission-b*, the story in which Bodyguard "causes" Victim's death by withholding the antidote strikes us as incomplete in the absence of any mention of the poison. In this case, the corresponding causal network is not self-contained.

7. The Main Idea

The main idea that I will defend is that counterfactual dependence is both necessary and sufficient for token causation in self-contained networks. Let us state this carefully:

TC Let <**V**, **E**> be a causal model, let X, $Y \in \mathbf{V}$, and let $X = x$ and $Y = y$. If the causal network connecting X to Y in <**V**, **E**> is self-contained, then $X = x$ is a token cause of $Y = y$ in <**V**, **E**> if and only if Y counterfactually depends upon X in <**V**, **E**>.

TC tells us that counterexamples to the simple counterfactual theory of causation arise only in causal networks that are not self-contained. The intuition behind *TC* is that when a causal network is self-contained, then whatever pattern of counterfactual dependence exists is due to the intrinsic nature of the causal relationship between the two events involved. By contrast, when a causal network is not self-contained, then the presence or absence of counterfactual dependence may not be diagnostic of the underlying causal relationship. When a causal network is not self-contained, that is an indication that some factor extrinsic to the system is interfering with the normal pattern of counterfactual dependence that would emerge from the underlying causal relationship (or lack thereof) between the two events.

TC is a bold claim, and no sooner is it stated than it must be qualified in a number of respects. First, if the variables X and Y are nonbinary, then it may turn out that some hypothetical changes in the value of X lead to different values of Y, while others do not; or that hypothetical changes in the value of X lead only to very small changes in the value of Y. This raises the question of how much and what kind of counterfactual dependence is needed for $X = x$ to count as a cause of $Y = y$. The discussion of section 10 below bears on this question, as do some of my own views about the role of contrast in causal claims,[11] but I will offer no comprehensive answer to this question here.

Second, *TC* imposes conditions on the relation of token causation in a model. In order for the events represented by $X = x$ and $Y = y$ to stand in the relation of token causation, the model must be an appropriate one.

Third, given the diversity of intuitions about token causation, and given my generally pragmatic orientation, I doubt that *TC* can survive in its absolute form.[12] The most I am prepared to say with any certainty is that we are strongly inclined to align our judgments of token causation with *TC*. That is, when we have counterfactual dependence in a self-contained network, we are strongly inclined to judge that there is token

11. See, for example, Hitchcock 1996 and references therein.
12. We will see one difficult kind of case in note 28.

causation; and when we have a self-contained network without counterfactual dependence, then we are strongly inclined to judge that there is no token causation. I leave it open whether these strong inclinations can sometimes be overridden by other considerations.

The reader cannot help but have noticed that *TC* provides a necessary condition for causation and a sufficient condition, but leaves a gap between them. What are we to say about cases where the causal network is not self-contained? I will not provide a complete answer to this question. In some cases, the silence of *TC* reflects a lack of consensus about the causal status of the case in question, but this will not always be the case. But even when *TC* issues no clear verdict, the notion of a self-contained network succeeds in capturing certain features of our judgments about the relevant cases. Moreover, *TC* can serve to distinguish certain cases where we do have clear intuitions from superficially similar cases where we have the opposite intuitions. We will see how *TC* functions by parading through the standard litany of test cases.[13]

8. Omission, Prevention, and Parasitic Dependence

All of the pieces needed for our treatment of omission and prevention are in place. In *Omission*, Assassin poisons Victim's coffee, while Bodyguard refrains from administering the antidote. What does our central principle *TC* say about the causal contribution of the two agents? The causal model of *Omission* is given in *Om*. The causal network connecting A to D is $\{A, D\}$. This network is self-contained, and the value of D counterfactually depends upon the value of A. Hence, *TC* rules that $A = 1$ is a token cause of $D = 1$: Assassin's poisoning the coffee is a token cause of Victim's death. The causal network connecting B to D is $\{B, D\}$. While the value of D depends counterfactually upon the value of B, the causal network is not self-contained, and *TC* delivers no definitive pronouncement.

13. There are two well-known types of case that I will not discuss. The first is "trumping" preemption (see Schaffer 2000b). I follow McDermott (2002) and Halpern and Pearl (2005) in thinking that trumping is a species of overdetermination and not of preemption. The second type of case I shall ignore is "preemptive prevention" (see Collins 2000). These cases should receive the same treatment as cases of early preemption (see section 11), with the following caveat: When we regard the possibility in which the backup preventer fails to be too remote to be considered seriously, we may simply omit the relevant variable from the model altogether. In the resulting model, *TC* will rule that there is no causation.

If we tweak the example slightly so that Bodyguard administers the antidote, thus saving Victim's life, we will have a case of prevention (called, naturally enough, *'Prevention'*). The causal model is as follows.[14]

Pr $A = 1$
$B = 1$
$D = A \,\&\, \sim B$

The causal graph is the same as that for *Om*, shown in figure 2. Again, the two causal networks are $\{A, D\}$ and $\{B, D\}$. The first is self-contained since $D = 0$ when $A = 0$. Since D does not counterfactually depend upon A, *TC* rules that Assassin's poisoning the coffee is not a token cause of Victim's survival. D does counterfactually depend upon B, but as in *Om*, the causal network connecting B to D is not self-contained: when B takes the value 0, D takes the value 1. Once again, *TC* delivers no definitive announcement.

Interested readers can convince themselves that cases of *double prevention* (Hall 2004) yield a similar result: counterfactual dependence in a causal network that is not self-contained.

TC is silent on the issue of whether omissions and preventions are genuine token causes. This silence is perhaps appropriate, given the diversity of opinion on the issue. Nonetheless, the general framework that I am advancing does seem to capture an important feature of the way we think about cases of omission and prevention. In *Omission*, for example, we straightforwardly regard Assassin's poisoning of the coffee as a token cause of Victim's death. Whether we countenance Bodyguard's inaction as a genuine cause or not, we recognize that the dependence of Victim's death upon Bodyguard's inaction is parasitic upon a primary causal process initiated by Assassin's action. This difference is captured by the difference between a causal network that is self-contained and one that is not. In cases where we have counterfactual dependence in a causal network that is not self-contained, then, I will say that we have *parasitic dependence*. Whether we regard *parasitic dependence* as genuine token causation or not is a matter of brute intuition, and I will follow *TC* in offering no pronouncement on this issue.

14. This assumes that Bodyguard administers the antidote independently of Assassin's action. If, instead, Bodyguard administers the antidote in response to the assassination attempt, the causal structure will be that given by model *CT*. This will not affect the relationship between Bodyguard's action and Victim's survival, although it could potentially affect the relationship between Assassin's action and Victim's survival. See section 11 for further discussion.

9. Causes versus Background Conditions

It is not uncommon to distinguish between causes and background conditions.[15] For example, we might count the careless tossing of the cigarette as a cause of the forest fire and the presence of oxygen in the atmosphere as a mere background condition. *TC* readily marks this distinction. The key is that a background condition corresponds to the default value of an appropriate variable. Consider the following example:

> *Background:* Assassin puts poison in Victim's coffee. Victim drinks the poison and dies. The poison would not have killed Victim if it were not for ordinary digestive processes that allowed the poison to enter Victim's bloodstream.

We may represent this scenario as follows:

> $A = 1$ if Assassin poisons Victim's coffee, 0 if not.
> $B = 0$ if Victim's digestive system is functioning properly, 1 if not.
> $D = 1$ if Victim dies, 0 if not.

The corresponding equations are:[16]

$$\begin{aligned} BG \quad & A = 1 \\ & B = 0 \\ & D = A \,\&\, {\sim}B \end{aligned}$$

The graph is the same as the one for *Omission*, shown in figure 2. Note that the default value of B corresponds to the normal functioning of Victim's digestive system—this is an ongoing biological process that will continue under its own steam in the absence of any interfering causes. In this case, the default outcome does not correspond to an absence or omission. Note that '$\sim B$' in the third equation refers to B taking the value 0, that is, to the occurrence of the relevant processes. *BG* has exactly the same structure as *Om*, hence *TC* rules that Assassin's poisoning the coffee is a straightforward cause of death, whereas the relationship between Victim's digestive processes and death is one of parasitic dependence.

15. See, for example, Hart and Honoré 1985, chap. 3, sec. 2. On the other hand, Mill (1843), Lewis (1973), and Hall (2004) all warn against such "invidious discrimination" (Lewis 1973, 559).

16. For simplicity, I have assumed that the failure of the digestive processes is not fatal—at least not within the time frame of the example. This does not fundamentally change the structure of the example.

TC does not rule on the question of whether background conditions are genuine causes or not, but it does capture the intuitive difference between background conditions and paradigmatic causes.

10. Causing and Affecting

It is common to make a distinction between causing some outcome to occur and merely affecting the outcome (see especially Mellor 1995, chap. 12). Consider the following example:

> *Affecting*: Assassin puts poison in Victim's coffee. The poison, if not counteracted, causes a painful death. Bodyguard puts a weak antidote in Victim's coffee. The antidote is not strong enough to neutralize the poison and hence save Victim's life, but at least it is strong enough to render Victim's death painless.

Here we would not say that Bodyguard's intervention caused Victim's death, or even that it caused Victim's painless death; rather we would say that Bodyguard's action affected Victim's death, or that it affected the manner in which the death occurred, or perhaps that it caused the death to be a painless one. By contrast, we have no trouble counting Assassin's action as a cause of Victim's death. What is the underlying distinction here?

We can capture some features of this example using a contrastive approach: Assassin's poisoning the coffee caused Victim to die painlessly rather than survive, but not to die painlessly rather than painfully; Backup's administering the antidote caused Victim to die painlessly rather than painfully, but did not cause Victim to die painlessly rather than survive. This is correct, but it does not seem to go far enough. Why do we judge Assassin's action to be *more of a cause* than Bodyguard's?

Let us model *Affecting*. As variables we choose:

> A = 1 if Assassin poisons Victim's coffee, 0 if not.
> B = 1 if Bodyguard administers the antidote, 0 if not.
> D = 2 if Victim dies painlessly, 1 if Victim dies painfully, 0 if he does not die.
> Def (D) = 0, Dev (D) = {1, 2}

Note that D is a ternary variable; it allows us to represent, not merely whether Victim dies, but also whether Victim's death is a painless or a

painful one. This variable has only one default value—corresponding to Victim's survival—but two possible deviant values. Our equations are:[17]

$$\begin{aligned} Af \quad & A = 1 \\ & B = 1 \\ & D = A(B + 1) \end{aligned}$$

The third equation tells us that D will take the value 0 when A does, and that when $A = 1$, then D will be equal to 1 or 2, depending upon the value of B. The causal graph is the same as that for *Om*, shown in figure 2.

Note that the value of D depends counterfactually upon the values of both A and B. If A had taken the value 0, then D would have been 0; whereas if B had taken the value 0, then D would have taken the value 1. The causal network connecting A to D is $\{A, D\}$. This network is self-contained: A satisfies *PSR* trivially, and D takes the default value 0 when its parent A does. Thus *TC* rules unambiguously that $A = 1$ is a token cause of $D = 2$. $\{B, D\}$, the causal network connecting B to D, is not self-contained: when B takes the default value 0, D takes the deviant value 1. The relationship between B and D is therefore one of parasitic dependence. Bodyguard's action can affect the manner of Victim's death only in virtue of a primary process that causes death to occur in the first place. Again, I will not attempt to rule on whether affecting involves genuine causation, but will satisfy myself with explaining the intuitive difference between affecting and causing.

11. Early Preemption, Transitivity, and the Counterexamples

Let us now turn to cases of early preemption, as illustrated by *Early Preemption* in section 3. Before looking at how *TC* treats these cases, however, it is worth briefly reviewing some other attempts to handle early preemption.

The best-known treatment of early preemption is that of Lewis (1973). Lewis's account presupposes that causation is transitive: if c causes d, which in turn causes e, then c is a cause of e. In *Early Preemption*, there will be an intermediate event between Assassin's poisoning of the coffee and Victim's death—for example, the presence of poison in the coffee shortly after Assassin poured it in. This event counterfactually depends

17. As with *Prevention*, we assume that Bodyguard acts independently and is not caused to add the antidote in response to Assassin's poisoning the coffee. If we change the story so that Bodyguard does act in response to Assassin's intervention, this will complicate the model slightly, but it will not affect the central features of the analysis.

upon Assassin's action: if Assassin had not poisoned the coffee, the poison would not have been present in the coffee at that time (although there would have been poison in the coffee later, after Backup put some in). Moreover, Victim's death counterfactually depends upon this event: if there had not been poison in the coffee, Victim would not have died. (Note that counterfactuals do not backtrack, so if there had been no poison in the coffee, Assassin still would have poisoned it and Backup would have refrained from doing so; the counterfactual requires us to imagine that the poison miraculously disappears.) The role of the interpolated event can readily be modeled by adding a suitable variable to *EP*; the details are left to the reader.[18] Since, on Lewis's account, counterfactual dependence is sufficient for causation,[19] we can conclude that Assassin's action caused the poison to be present, which in turn caused Victim's death. By transitivity, Assassin's poisoning the coffee caused Victim's death.

Unfortunately, causation does not seem to be transitive in general. The following sort of case is often taken to be a counterexample:

> *Counterexample to Transitivity:* Assassin puts poison in the coffee. Bodyguard responds by putting antidote in the coffee, which neutralizes the poison. (The antidote is harmless when taken alone.) Victim drinks the coffee and survives. If Assassin hadn't poisoned the coffee, Bodyguard would not have administered the antidote. If Bodyguard hadn't administered the antidote, Victim would have died from the poison.

Assassin's poisoning the coffee caused Bodyguard to administer the antidote, and Bodyguard's administering the antidote caused Victim to survive, but most people judge that Assassin's poisoning the coffee is not a token cause of Victim's survival.

There are a couple of different ways in which one might attempt to block this counterexample. First, one might reject the second causal claim—that Bodyguard's administering the antidote caused Victim to survive—on the grounds that this is a case of prevention rather than genuine token causation. I have three replies. First, this response is not available to Lewis since he takes prevention to be genuine causation. Second,

18. See Hitchcock 2001, secs. 4 and 6, for discussion of a similar example.

19. As we saw in the previous section, the present account does not endorse this principle; however, in this example the relevant causal networks are self-contained, so the sufficiency claim will hold here.

one can construct more complex counterexamples to transitivity, where none of the steps in the causal chain involve parasitic causation.[20] Third, even if we deny that prevention is genuine causation, we might wonder why we do not judge that Assassin's action prevented Victim's death, given that Assassin's action caused something that did prevent Victim's death. So even if this is not a genuine counterexample to the transitivity of token causation *per se*, it would still be interesting to explain how this case differs from paradigmatic cases of prevention.

Lewis himself rejects this sort of counterexample by denying the last claim—that Assassin's poisoning the coffee is not a cause of Victim's survival (see, for example, Lewis 2000, sec. 8). I will not attempt any rebuttal of Lewis's claim here. It seems clear enough that we are *more strongly inclined* to judge that there is token causation in *Early Preemption* than we are in *Counterexample to Transitivity*, and I will content myself to explain this difference.

Let us construct a causal model of *Counterexample to Transitivity*. The variables will be as follows:

A = 1 if Assassin poisons the coffee, 0 if not.
B = 1 if Bodyguard administers the antidote, 0 if not.
D = 1 if Victim dies, 0 if not.

The corresponding equations are:

CT A = 1
B = A
$D = A \,\&\, \sim B$

The causal graph for *CT* is the same as the one for *EP*, shown in figure 1.

In Hitchcock 2001, I offered my own account of early preemption. If we look at the causal graph of *EP*, shown in figure 1, we can see that there are two paths from A to D: one is direct, and the other runs through B. Intuitively, it is in virtue of the first path that $A = 1$ is a token cause of $D = 1$; nonetheless, the value of D fails to depend counterfactually upon the value of A because there is some sort of cancellation along these two

20. For example, suppose that Assassin has poisoned Victim with a single dose of poison. This prompts Bodyguard to respond by administering an antidote that is just strong enough to neutralize the poison. Backup, in turn, responds by adding additional poison to the coffee, which is not neutralized. Victim drinks the coffee and dies. Here, Bodyguard's adding the antidote caused Backup to add more poison, which in turn caused Victim to die. All of these are deviant outcomes. Yet still, most would judge that Bodyguard's administering the antidote did not cause Victim to die.

paths. To show that $A = 1$ is a token cause of $D = 1$, we need to isolate the influence of the former on the latter along the direct path. We can do this by "freezing" the indirect path. That is, when we hold the value of B fixed at its actual value of 0, the counterfactual dependence of D upon A is restored. In effect, what we are doing is evaluating the counterfactual: "If Assassin had not put poison in the coffee, and Backup (still) did not put poison in the coffee, then Victim would not have died." This counterfactual is intuitively true, and *EP* reproduces this verdict when we substitute the equations for A and B with $A = 0$ and $B = 0$, respectively.

More generally, the proposal is:

> *AP* Let <**V**, **E**> be a causal model, and let $X, Y \in V$. Let P = $\{X, \ldots, Y\}$ be a path from X to Y in <**V**, **E**>. **P** is *active* in the causal model <**V**, **E**> if and only if Y depends counterfactually upon X within the new system of equations **E'** constructed from **E** as follows: for all $Z \in$ **V**, if Z does not belong to **P**, then replace the equation for Z with a new equation that sets Z equal to its actual value in **E**. (If there are no variables that do not belong to **P**, then **E'** is just **E**.) Then $X = x$ is a token cause of $Y = y$ just in case there is an active path from X to Y.[21]

Several other accounts, including those of Halpern and Pearl 2001, 2005 and Yablo 2002, 2004, treat *Early Preemption* in essentially the same way.

Unfortunately, this account of early preemption runs into problems with a new sort of case:[22]

> *Counterexample to Hitchcock.* Assistant Bodyguard puts a harmless antidote in Victim's coffee. Buddy then poisons the coffee, using a type of poison that is normally lethal, but which is countered by the antidote. Buddy would not have poisoned the coffee if Assistant had not administered the antidote first. Victim drinks the coffee and survives.[23]

21. Note that in Hitchcock 2001, I did not require that all variables in V\P be held fixed, only those that lie on alternate paths from X to Y. This is more economical, but makes no difference to whether a route is active or not. A more adequate account would weaken *AP* to allow for active *routes*, where a route comprises multiple paths from X to Y; we will not concern ourselves with this modification here.

22. It will also run into trouble in *Counterexample to Transitivity* if our causal model includes a new variable interpolated between A and D. See Hitchcock 2001, sec. 9; Hitchcock n.d., secs. 4, 5.

23. Examples with a similar structure have been independently suggested to me by

In order to understand this scenario, we might suppose that Assistant Bodyguard is up for a promotion (the bodyguard equivalent of tenure) and wants to make it look as though he has foiled an assassination attempt. Buddy is helping him. They don't want to actually harm Victim, so Buddy makes certain that Assistant has added the antidote before putting poison in the coffee.

Here is the model:

CH $A = 1$ if Assistant Bodyguard administers the antidote, 0 if not.
$B = 1$ if Buddy poisons the coffee, 0 if not.
$D = 1$ if Victim dies, 0 if not.

$A = 1$
$B = A$
$D = \sim A \ \& \ B$

The causal graph is the same as in figure 1.

Many people, but by no means all, have the intuition that Assistant's adding the antidote to the coffee is not a cause of Victim's survival. Some respondents have the opposite intuition, and many have no clear intuitions at all. The account of Hitchcock 2001 clearly rules that Assistant's action is a cause of Victim's survival. The path $\{A, D\}$ is active, as can be seen by holding fixed the variable B at its actual value of 1. Given that Buddy poisoned the coffee, Victim would not have survived if Assistant had not administered the antidote. The accounts of Halpern and Pearl 2001, 2005 and Yablo 2002, 2004 yield the same verdict.[24] Given the diversity of opinion on this example, the counterexample can hardly be seen as crushing; still, it would be nice to have an account of how *Counterexample to Hitchcock* differs from *Early Preemption*.

In each of the three causal models, *EP*, *CT*, and *CH*, the causal network connecting A to D is $\{A, B, D\}$. As we noted in section 6, this network is not self-contained in *EP*: B takes the deviant value 1 when its lone parent, A, takes the default value 0. By contrast, the causal network $\{A, B,$

Michael McDermott (personal communication) and Gunnar Björnsson (2007). I offer another one in Hitchcock 2003. The example presented here (which also appears in Hitchcock n.d.) is by far the cleanest of the lot.

24. Yablo's case is tricky since his theory does not treat the existence of an active route as sufficient for causation. Rather than delve into the details of his account, I will satisfy myself with the claim that *Counterexample to Hitchcock* poses a prima facie difficulty for his account.

D} is self-contained in both *CT* and *CH*. In particular, in both models, *B* takes the deviant value of 1 only when *A* does. Since there is no counterfactual dependence of *D* upon *A* in any of the models, *TC* rules that $A = 1$ is not a token cause of $D = 1$ in *CT* or *CH*, but it leaves the causal status of *EP* unsettled. Unlike both *Early Preemption* and *Prevention*, *Counterexample to Transitivity* and *Counterexample to Hitchcock* belong to the class of cases where *TC* delivers a clear negative verdict.

It is perhaps worth trying to articulate informally the work that *TC* is doing in these cases. In *Counterexample to Transitivity* and *Counterexample to Hitchcock*, it is natural to say something like the following: In each case, the putative cause (Assassin's poisoning the coffee in *Counterexample to Transitivity*, and Assistant's putting antidote in the coffee in *Counterexample to Hitchcock*) introduces a threat to Victim's life (by causing poison to be present in the coffee). At the same time, the putative cause also counters that threat (by causing antidote to be added to the coffee). Because a threat to Victim's life is countered, we might be inclined to think of *Counterexample to Transitivity* and *Counterexample to Hitchcock* as cases of prevention (whether or not we take prevention to be genuine causation). But these cases are different from genuine cases of prevention; in genuine cases of prevention, the preventer is independent of the threat countered. In *Counterexample to Transitivity* and *Counterexample to Hitchcock*, by contrast, the putative preventer is itself the source of the threat that is countered. The putative cause gives with one hand and takes away with the other.[25] The idea that there is no *independent* threat to Victim's life that is being neutralized is captured by the notion of a self-contained network.

Note that *TC* by itself doesn't explain our strong inclination to judge that Assassin's poisoning the coffee *is* a token cause of Victim's death in *Early Preemption*, although it does explain why we lack a strong inclination to deny this claim. It follows that *TC* cannot be the whole story on token causation. Indeed, *TC*, by itself, delivers a verdict only in cases where the causal network is self-contained. In cases where the causal network is not self-contained and there is counterfactual dependence, we have parasitic dependence, which one may countenance as genuine causation or not, according to one's taste. That leaves only cases where we have no counterfactual dependence in a causal network that is not self-contained. *Early Preemption* falls into this final category. We may complete the analysis by adopting the account of Hitchcock 2001 or Halpern and Pearl 2001, 2005 and restricting it to this final case. An

25. See the similar comments in Yablo 2004, sec. 5.

alternate picture that respects the divergence of opinion in cases like *Counterexample to Hitchcock* is that *TC* and the active path account both capture important features of the way we think about causation, and that when they conflict, intuitions will be divided.[26]

12. Symmetric Overdetermination

We turn next to cases of *symmetric overdetermination*. Here is a standard type of example:

> *Symmetric Overdetermination.* Assassin and Badgirl simultaneously poison Victim's coffee with identical doses of a lethal poison. Either dose by itself would have sufficed for Victim's death. Victim drinks the coffee and dies. He would have survived if the coffee had not been poisoned.

The model is straightforward. The variables have their usual interpretation, and the equations are:

$$\begin{array}{ll} SO & A = 1 \\ & B = 1 \\ & D = A \vee B \end{array}$$

The graph conforms to the pattern of figure 2. Victim's death does not depend counterfactually upon the actions of either agent. Most, but not all,[27] have the intuition that each assassin's poisoning of the coffee counts as a token cause of Victim's death.

This case has some features in common with *Early Preemption*: Victim's death would have been counterfactually dependent upon Assassin's action if Badgirl had not been there. Unfortunately, the treatment of early preemption canvassed above will not help us here: even if we hold B fixed at its actual value of 1, we cannot restore the counterfactual dependence of D upon A.

On the other hand, if we fix the value of B at 0, then we do have counterfactual dependence of D upon A. This suggests that *AP* will need to be weakened: sometimes in order to reveal an active path it will be necessary hold the off-path variables fixed at nonactual values. But we cannot be too permissive; we cannot allow the off-path variables to be fixed at arbitrary values. Consider *Prevention*, for example. If *Bodyguard* had not

26. See Hitchcock n.d. for further development of this idea.
27. Lewis (1973) claims to have no clear intuitions about this case, for example.

put antidote in the coffee, then whether or not Victim died would have depended counterfactually upon whether or not Assassin poisoned the coffee. But that hardly makes Assassin's poisoning of the coffee a cause of Victim's survival. What is needed is a way of restricting the values to which off-path variables may be set. In Hitchcock 2001 I proposed that an off-path variable X may be set to a value x if doing so does not change the value of any of the variables that lie on the path. Thus in *SO*, changing the value of B from 1 to 0 makes no difference to the value of either A or D. In *Pr*, by contrast, changing the value of B from 1 to 0 does change the value of D, so this would not be a permissible setting of the value B. Halpern and Pearl (2001, 2005) show that this condition is too restrictive in general and provide a complex alternative. Fortunately, the details are not essential here: the important point is that any treatment of *Symmetric Overdetermination* in the tradition of Hitchcock 2001 or Halpern and Pearl 2001, 2005 is going to require us to allow $A = 1$ to count as a cause of $D = 1$ on the grounds that the latter would counterfactually depend upon the former if B were set to 0.

Hiddleston (2005, 31) has offered a counterexample to any such approach.

> *Bogus Prevention.* Assassin is in possession of a lethal poison, but has a last minute change of heart and refrains from putting it in Victim's coffee. Bodyguard puts antidote in the coffee, which would have neutralized the poison had there been any. Victim drinks the coffee and survives.

BP $A = 0$
$B = 1$
$D = A \,\&\, \sim B$

The graph is that shown in figure 2. Note that setting A to 1 will not change the value of the other variables. However, if we set A to 1, D will counterfactually depend upon B. Thus the accounts of Hitchcock 2001 and of Halpern and Pearl 2001, 2005 will rule that $B = 1$ is a cause of $D = 0$. But it would seem very odd to say that Bodyguard's administration of the antidote prevented Victim's death, or that it was an overdetermining cause of Victim's survival. Victim's life was never at risk in the first place.

As always, I won't attempt to legislate on the correct intuitions on these cases. It seems clear that we are more strongly inclined to deny causation in *Bogus Prevention* than in *Symmetric Overdetermination*, and it is this feature of our intuitions that *TC* can capture. In *BP*, the causal

network connecting B to D is $\{B, D\}$. This network is self-contained: B satisfies *PSR* trivially, and D takes the value 0 when B does. Since D does not counterfactually depend upon B, *TC* rules that $B = 1$ is not a token cause of $D = 0$. In *SO*, the causal network connecting A to D is $\{A, D\}$. This network is not self-contained: D takes the deviant value 1 when A takes the default value 0. Thus, even though D does not depend counterfactually upon A, *TC* does not exclude $A = 1$ as a token cause of $D = 1$. We may then appeal to the account of Hitchcock 2001 or Halpern and Pearl 2001, 2005 to rule on this case.[28]

13. Late Preemption

Cases of late preemption have been among the most recalcitrant for counterfactual theories of causation. (In particular, Lewis's strategy for dealing with early preemption will not work for these cases.) Here is an example that we will work with:

> *Late Preemption*: Assassin and Badgirl both put poison in Victim's coffee, which he drinks at noon. Assassin uses a fast-acting poison that kills within one hour. Badgirl uses a slow-acting poison that takes from one to two hours to kill. At 1:00 p.m., Victim is dead. If only Badgirl had administered the poison, Victim would have been alive at 1:00 p.m., but dead by 2:00 p.m.

28. An anonymous referee noted that this treatment of symmetric overdetermination is problematic in the case of symmetrically overdetermined prevention. Suppose that Victim's coffee has been poisoned and that both Bodyguard and Assistant put antidote in the coffee, where each dose of antidote is sufficient to neutralize the poison. Readers may confirm that the relevant causal networks will be self-contained, and hence *TC* will rule that neither Bodyguard nor Assistant caused Victim to remain alive. This may seem counterintuitive, although I think that we are less inclined to judge that there is token causation in this case than in *Symmetric Overdetermination*. Thus it may be possible to handle this case along the lines suggested at the end of section 1: when *TC* conflicts with the accounts of Hitchcock 2001 or Halpern and Pearl 2001, 2005, our intuitions will be conflicted. Note also that there is an asymmetry between this sort of case and *Symmetric Overdetermination*. In the latter case, it does not matter which assassin puts the poison in first: Victim drinks both doses of poison, and both contribute to his death. In the case where both bodyguards administer antidotes, it seems that only the first to put her antidote into the coffee causes the coffee to be neutralized. Thus, the structure seems to be more like a case of late preemption than symmetric overdetermination. Indeed, if we model the scenario with additional variables representing the state of the coffee right after the actions of the two bodyguards, we will get a structure similar to *LP*. In this case, *TC* will rule that only the first bodyguard to administer the antidote caused Victim to survive. The proof is left as an exercise for the reader.

Intuitively, Assassin's poisoning of the coffee caused Victim's death, whereas Badgirl's did not; but Victim's death is counterfactually dependent upon neither. This case turns out to be rather tricky to model. We may start off in the usual fashion, with two of the relevant variables:

A = 1 if Assassin poisons the coffee, 0 if not.
B = 1 if Badgirl poisons the coffee, 0 if not.

The problem remains of how to represent the various possibilities surrounding Victim's death. It clearly will not do to have a single binary variable, D, that takes the value 1 if Victim dies and 0 if he survives. This would clearly yield a causal model identical to *SO* and would obscure the difference in causal status between the two acts of poisoning. One natural suggestion would be to use a ternary variable, much as we did in modeling *Affecting*. We could let the variable D take the value 1 if Victim dies between noon and 1:00 p.m., 2 if Victim dies between 1:00 and 2:00 p.m., 0 if Victim does not die at all. Unfortunately, such a variable would violate one of the strictures on causal modeling discussed in section 3. The values of this variable do not represent possibilities that are incompatible on broadly logical or conceptual grounds: Victim's death before 1:00 *prevents* him from dying between 1:00 and 2:00 p.m. This "causal" relationship is obscured if we use the ternary variable suggested. Instead we must use two variables:

D_1 = 1 if Victim is dead at 1:00 p.m., 0 otherwise.
D_2 = 1 if Victim is dead at 2:00 p.m., 0 otherwise.

This raises an interesting question: just what, in the model, represents the event "Victim's death"? One possibility would be to say that this event corresponds to the disjunction $D_1 = 1$ or $D_2 = 1$. In any event, I think that we will have adequately captured our intuitions about the case if we can show that $A = 1$ is a cause of both $D_1 = 1$ and $D_2 = 1$, while $B = 1$ is a cause of neither.

The assignment of defaults is also tricky in this case. The default values of A, B, and D_1 are all 0, as usual. D_2, however, is an inertial variable. If Victim is still alive at 1:00 p.m., then the default is that he will survive until 2:00; but if Victim is already dead by 1:00, then the default is that he will still be dead at 2:00. Therefore:

$\text{Def}(D_2) = D_1$

The equations are:

$$\begin{aligned} LP \quad & A = 1 \\ & B = 1 \\ & D_1 = A \\ & D_2 = B \vee D_1 \end{aligned}$$

The causal graph is shown in figure 3. The third equation says that Victim will be dead at 1:00 just in case Assassin poisons him, and the fourth equation says that Victim will be dead at 2:00 just in case he is dead at 1:00 or Badgirl poisons him.

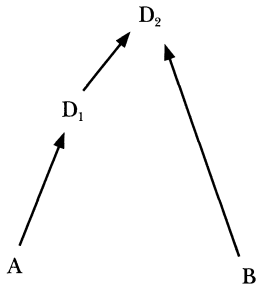

Figure 3.

Most treatments of late preemption focus on a similarity with cases of early preemption. In *Early Preemption*, there is a causal process that would have killed Victim had Assassin not poisoned his coffee. But that process is in fact cut short; in our example, Backup never puts the poison in the coffee. Similarly, in *Late Preemption*, let us suppose that there are certain physiological processes that take place shortly before death whenever the slow-acting poison causes death. These processes would have taken place within Victim's body some time after 1:00 p.m. if Assassin hadn't administered the fast-acting poison. These processes did not take place, however, because Victim was already dead at the time. In both cases, the difference between the successful cause and the preempted backup was that the causal process initiated by the former was allowed to run to completion, whereas the causal process initiated by the latter was not. In cases of early preemption, the backup process is cut off *before* the effect occurs, whereas in cases of late preemption, the process is cut off by the effect itself.

In my opinion, the most successful treatment of late preemption that employs this kind of strategy is that of Halpern and Pearl (2005, sec. 4.2).[29] Let us add a variable representing the physiological processes typically induced by the slow-acting poison.

29. The treatment of Pearl 2000, sec. 10.3.5 is similar, although that of Pearl and Halpern 2001 is flawed, for reasons discussed in Pearl and Halpern 2005.

$P = 1$ if these processes occur some time after 1:00, 0 if they do not occur at this time.

The new system of equations is:

LP' $A = 1$
$B = 1$
$D_1 = A$
$P = B \,\&\, \sim D_1$
$D_2 = P \vee D_1$

The graph is shown in figure 4. It should be uncontroversial that $A = 1$ is a cause of $D_1 = 1$, whereas $B = 1$ is not, so we will focus on the effect $D_2 = 1$. The reader can convince herself that $\{A, D_1, D_2\}$ is an active route: holding fixed that $P = 0$, D_2 counterfactually depends upon A. By contrast, the route $\{B, P, D_2\}$ is not active. Nor can the strategy employed for handling symmetric overdetermination be used here since changing the value of D_1, which would be necessary to render D_2 counterfactually dependent upon B, would change the value of P. Thus the account of Hitchcock 2001 and Halpern and Pearl 2001, 2005, when applied to LP', yield the intuitively correct result that $A = 1$ is a cause of $D_2 = 1$ and that $B = 1$ is not.

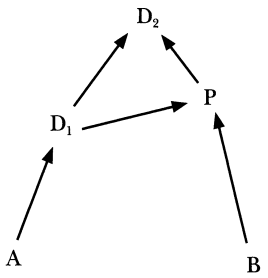

Figure 4.

This analysis is not unproblematic, however. First, it seems to involve *oversophistication*. We really have no idea what these physiological processes are, they were never mentioned in the original story, and it seems implausible that we are reasoning about them when we judge that Assassin's poisoning the coffee killed Victim, whereas Badgirl's action did not. There is a second, more subtle problem with the model LP'.[30] According to the fourth equation, P may come to take the value 0 in two very different ways. The processes might fail to occur because Victim is healthy, with none of the slow-acting poison in his body, but they might fail to occur because Victim is dead. The value $P = 0$ does not distinguish between these two possibilities. But surely these two states have very different implications about whether or not Victim is alive at 2:00.

30. This is essentially the problem raised by Hall and Paul 2003, sec. 6.

The fifth equation tells us that if Victim had been alive at 1:00, and if these processes did not occur shortly after 1:00, then Victim would not be dead at 2:00. But is this counterfactual clearly true? Couldn't the second conjunct of the antecedent be realized by having Victim die shortly after 1:00? Indeed, the state represented by $P = 0$ was in fact realized by Victim's being dead at the appropriate time, so why not assume that's how it would be realized in the relevant counterfactual scenario?

I do not take these objections to be knockdown, but they suffice to motivate interest in an alternative approach. I will use the original model *LP*, which has the virtue of being simpler than the extended model *LP′*.[31] As figure 3 suggests, my treatment of *Late Preemption* will be closer to the treatment of *Symmetric Overdetermination* than to that of *Early Preemption*. It is easy to see how an adequate account of symmetric overdetermination will yield the verdict that Assassin's action is a cause of Victim's being dead at 2:00. We allow the variable B to be held fixed at the nonactual value 0, and counterfactual dependence of D_2 upon A is restored. The problem is how to exclude Badgirl's poisoning the coffee as a cause of Victim's being dead at 2:00. Intuitively, my suggestion is that what excludes Badgirl's action as a cause is not that the relevant process is cut short, but rather that an important precondition for its causal efficacy is no longer present: you just can't kill someone who is already dead!

First, note that *TC* clearly rules that $A = 1$ is a cause of $D_1 = 1$, whereas $B = 1$ is not. The causal network $\{A, D_1\}$ is self-contained, and D_1 depends counterfactually on A; $\{B, D_1\}$ is trivially self-contained, and D_1 does not depend upon B. So now let us look at the causal networks connecting A and B to D_2. The network $\{B, D_2\}$ is self-contained. Whatever value B takes, D_2 will take the value 1, which is its default value, given that D_1 also takes the value 1. D_2 does not depend counterfactually upon B, so *TC* rules that $B = 1$ is not a token cause of $D_2 = 1$. By contrast, the causal network $\{A, D_1, D_2\}$ connecting A to D_2 is not self-contained. When D_1 takes the default value 0, D_2 takes the value 1. This is now a *deviant* value of D_2 since it differs from the value of D_1. Thus D_2 takes a deviant value when its parent, D_1, takes a default value. There is a sense in which both Assassin's poisoning the coffee and Badgirl's poisoning the coffee are overdeterminers of Victim's being dead at 2:00. But the way in which Victim's state at 1:00 affects the default value of Victim's state at 2:00 renders Badgirl's action a bogus overdeterminer of Victim's death, much

31. Note, however, that use of the extended model *LP′* would not undermine the present analysis.

as Bodyguard's action was a bogus overdeterminer of Victim's survival in *Bogus Prevention.*

14. Conclusion

I have, in effect, proposed a four-fold distinction. If we want to know whether c is a token cause of e, we must construct an appropriate causal model <**V**, **E**> in which these events are represented as the values of variables $X = x$ and $Y = y$. Within our causal model we ask: Does Y depend counterfactually upon X? Is the causal network connecting X and Y self-contained? There are four possible answers.

1. *Counterfactual dependence in a self-contained network.* In this case, we will feel strongly compelled to say that c is a token cause of e.

2. *No counterfactual dependence in a self-contained network.* In this case, we will feel strongly compelled to say that c is not a token cause of e.

3. *Counterfactual dependence in a network that is not self-contained.* These are cases of parasitic dependence. In these cases, there may well be fundamental disagreement about whether to call c a token cause of e, or to accord it some kind of subsidiary causal status. Even in these cases, the concept of a self-contained network does a good job in capturing the contours of our intuitions.

4. *No counterfactual dependence in a network that is not self-contained.* Cases of preemption and overdetermination fall into this category. In these cases, *TC* allows there to be token causation despite the absence of counterfactual dependence. *TC* by itself does not rule one way or the other in these cases. It is therefore necessary to invoke some further account, such as that of Hitchcock 2001 or Halpern and Pearl 2001, 2005 to pass judgment on these cases. Nonetheless, *TC* plays a valuable role in distinguishing cases of preemption and overdetermination from superficially similar cases where we judge that there is no token causation.

References

Barker, S. 2004. "Analyzing Chancy Causation without Appeal to Chance-Raising." In Dowe and Noordhof 2004, 120–37.

Beebee, H. 2004. "Causing and Nothingness." In Collins, Hall, and Paul 2004, 291–308.

Björnsson, G. 2007. "How Effects Depend on Their Causes, Why Causal Transitivity Fails, and Why We Care about Causation." *Philosophical Studies* 133: 349–90.

Campbell, J. K., M. O'Rourke, and H. S. Silverstein, eds. 2007. *Causation and Explanation.* Cambridge, MA: MIT Press.

Collins, J. 2000. "Preemptive Prevention." *Journal of Philosophy* 97: 223–34 (reprinted in Collins, Hall, and Paul 2004, 107–17).

Collins, J., N. Hall, and L. Paul, eds. 2004. *Causation and Counterfactuals.* Cambridge, MA: MIT Press.

Colton, R. n.d. "Preemption and the Counterfactual Analysis of Causation." Unpublished manuscript.

Dowe, P. 2000. *Physical Causation.* Cambridge: Cambridge University Press.

Dowe, P., and P. Noordhof, eds. 2004. *Cause and Chance.* London and New York: Routledge.

Ganeri, J., P. Noordhof, and M. Ramachandran. 1996. "Counterfactuals and Preemptive Causation." *Analysis* 55: 219–25.

———. 1998. "For a (Revised) PCA-Analysis." *Analysis* 58: 45–47.

Glymour, C., and F. Wimberly. 2007. "Actual Causes and Thought Experiments." In Campbell, O'Rourke, and Silverstein 2007.

Hall, N. 2004. "Two Concepts of Causation." In Collins, Hall, and Paul 2004, 225–76.

Hall, N., and L. Paul. 2003. "Causation and Preemption." In *Philosophy of Science Today*, ed. P. Clark and K. Hawley, 100–130. Oxford: Oxford University Press.

Halpern, J., and J. Pearl. 2001. "Causes and Explanations: A Structural-Model Approach—Part 1: Causes." In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 194–202. San Francisco: Morgan Kaufmann.

———. 2005. "Causes and Explanations: A Structural-Model Approach—Part 1: Causes" (expanded version). *British Journal for the Philosophy of Science* 56: 843–87.

Hart, H. L. A., and T. Honoré. 1985. *Causation in the Law.* 2nd ed. Oxford: Oxford University Press.

Hiddleston, E. 2005. "Causal Powers." *British Journal for the Philosophy of Science* 56: 27–59.

Hitchcock, C. 1996. "Farewell to Binary Causation." *Canadian Journal of Philosophy* 26: 267–82.

———. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *Journal of Philosophy* 98: 273–99.

———. 2003. "Of Humean Bondage." *British Journal for the Philosophy of Science* 54: 1–25.

———. 2004a. "Do All and Only Causes Raise the Probabilities of Effects?" In Collins, Hall, and Paul 2004, 403–17.

———. 2004b. "Routes, Processes, and Chance-Lowering Causes." In Dowe and Noordhof 2004, 138–51.

———. 2007. "What's Wrong with Neuron Diagrams?" In Campbell, O'Rourke, and Shier 2007, 69–92.

———. n.d. "Token Causation." Unpublished manuscript.
Leibniz, G. 1956 [1716]. Third Letter to Samuel Clarke, February 25, 1716. In *The Leibniz-Clarke Correspondence*, ed. H. G. Alexander, 25–30. Manchester: Manchester University Press.
Lewis, D. 1973. "Causation." *Journal of Philosophy* 70: 556–67 (reprinted in Lewis 1986c, 159–72).
———. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13: 455–76 (reprinted in Lewis 1986c, 32–52).
———. 1986a. "Postscripts to 'Causation'." In Lewis 1986c, 159–213.
———. 1986b. "Events." In Lewis 1986c, 241–69.
———. 1986c. *Philosophical Papers*, vol. 2. Oxford: Oxford University Press.
———. 2000. "Causation as Influence." *Journal of Philosophy* 97: 182–97 (expanded version in Collins, Hall, and Paul 2004, 75–106).
Maudlin, T. 2004. "Causation, Counterfactuals, and the Third Factor." In Collins, Hall, and Paul 2004, 419–43.
McDermott, M. 1995. "Redundant Causation." *British Journal for the Philosophy of Science* 46: 523–44.
———. 2002. "Causation: Influence vs. Sufficiency." *Journal of Philosophy* 99: 84–101.
Mellor, D. H. 1995. *The Facts of Causation*. London: Routledge.
———. 2004. "For Facts as Causes and Effects." In Collins, Hall, and Paul 2004, 309–23.
Menzies, P. 2004a. "Difference-Making in Context." In Collins, Hall, and Paul 2004, 139–80.
———. 2004b. "Causal Models, Token Causation, and Processes." *Philosophy of Science* 71 (Proceedings): S820–S832.
Mill, J. S. 1843. *A System of Logic*. London: Parker and Son.
Noordhof, P. 1999. "Probabilistic Causation, Preemption, and Counterfactuals." *Mind* 108: 95–125.
———. 2004. "Prospects for a Counterfactual Theory of Causation." In Dowe and Noordhof 2004, 188–201.
Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Ramachandran, M. 1997. "A Counterfactual Analysis of Causation." *Mind* 106: 263–77.
———. 1998. "The M-Set Analysis of Causation." *Mind* 107: 465–71.
———. 2004a. "Indeterministic Causation and Varieties of Chance-Raising." In Dowe and Noordhof 2004, 152–62.
———. 2004b. "A Counterfactual Analysis of Indeterministic Causation." In Collins, Hall, and Paul 2004, 387–402.
Schaffer, J. 2000a. "Causation by Disconnection." *Philosophy of Science* 67: 285–300.

———. 2000b. "Trumping Preemption." *Journal of Philosophy* 97: 165–81 (reprinted in Collins, Hall, and Paul 2004, 59–73).
Woodward, J. 2003. *Making Things Happen.* Oxford: Oxford University Press.
Yablo, S. 2002. "De Facto Dependence." *Journal of Philosophy* 99: 130–48.
———. 2004. "Advertisement for a Sketch of an Outline of a Prototheory of Causation." In Collins, Hall, and Paul 2004, 119–37.