Counterfactual Theories

L. A. Paul

The Oxford Handbook of Causation
Edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies

Print Publication Date: Nov 2009
Subject: Philosophy, Metaphysics, Philosophy of Mathematics and Logic, Philosophy of Science
Online Publication Date: Jan 2010 DOI: 10.1093/oxfordhb/9780199279739.003.0009

Abstract and Keywords

Counterfactual analyses have received a good deal of attention in recent years, resulting in a host of counterexamples and objections to the simple analysis and its descendants. The counterexamples are often complex and can seem baroque to the outsider (indeed, even to the insider), and it may be tempting to dismiss them as irrelevant or uninteresting. But while we may be able to ignore some counterexamples because the intuitions they evoke are unclear or misguided, the importance of investigating the causal relation via investigating counterexamples should not be underestimated.

Keywords: counterfactual theories, counterexamples, intuitions, causal relation, counterfactual analysis, causation

I hit the eight ball into the corner pocket. If I hadn't hit the eight ball, it wouldn't have rolled into the corner pocket. Many think that the counterfactual 'if I hadn't hit the eight ball, it wouldn't have rolled into the corner pocket' captures something essential and fundamental about this instance of causation: my hitting the eight ball caused it to roll into the pocket, since if I hadn't hit it, it wouldn't have rolled into the pocket. Counterfactual analyses of causation seek to exploit this insight by constructing analyses of the causal relation (or our concept of it) in terms of counterfactual dependence. The central thought behind a counterfactual analysis of causation is that the relation of counterfactual dependence between *E*, the *eight ball rolling into the corner pocket* and *C*, *my hitting the eight ball* somehow captures the fact that there is a causal relation between these events. That is, *C* causes *E* because the counterfactual 'if not *C*, then not *E*' is true. To the extent that this is successful, we have a counterfactual analysis of causation.[1]

Counterfactuals are subjunctive conditionals of the form, 'if it were the case that *A*, then it would be the case that *B*'. Counterfactual analyses of causation focus on counterfactuals that tell us what would have been the case if the world had been different. In the main, they focus on counterfactuals concerning temporally successive, suitably distinct events *C* and *E* that describe cases where, if *C* had not (p. 159) occurred, *E* would not have occurred. Some counterfactual analyses are developed in terms of probabilistic counterfactuals, for example, if *C* had not occurred, *E* would not have had the probability of oc-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

curring that it did have. For reasons of expediency, I will focus on analyses built with deterministic conditionals.

Such counterfactuals describe relations of counterfactual dependence. Many philosophers take counterfactual dependence between (successive, suitably distinct) events to be sufficient for causation (Lewis 1973*b*). If event *E* would not occur if event *C* were not to occur, then *C* is a cause of *E*. However, since there are intuitively clear cases of causation without simple counterfactual dependence, counterfactual dependence is not necessary for causation: it is not the case that *C* is a cause of *E only* when *E* depends on *C*. *E* might still occur, otherwise caused—even if *C* is a cause of *E* in the actual world (see the discussions of pre-emption and overdetermination in sects. 3.2 and 3.3 below). Hence, the causal relation between *C* and *E* cannot be reduced to the relation of counterfactual dependence between *C* and *E*.

David Lewis, the most prominent contemporary defender of a counterfactual analysis, tries to solve this wrinkle by taking the ancestral of the counterfactual dependence relation. If we call the relation of the ancestral of counterfactual dependence '*R*', *C* stands in *R* to *E* iff *E* counterfactually depends on *C*, or *C* is connected by stepwise counterfactual dependence to *E*. Thus, even in cases where it is false that *E* depends directly on *C*, since if *C* had not occurred, *E* would have occurred (by being otherwise caused), as long as there is stepwise counterfactual dependence between *C* and *E*, *C* stands in *R* to *E*. A very simple and elegant account of causation could then take either dependence or stepwise dependence to be necessary for causation, and hold that *C* is a cause of *E* iff *C* stands in *R* to *E*. Call this the 'simple account'. David Lewis defended the simple account in his seminal (1973*b*) article, 'Causation' (reprinted in his 1986*a*). Defenders of similar accounts include Mackie (1965) and Lyon (1967).

Unfortunately, as Lewis himself quickly realized, the simple account is plagued with a host of problems. One obvious potential problem is that *R* is transitive, so if causation is the relation *R* then causation must also be transitive. Many, including Lewis, welcome this result, since they believe that the causal relation is transitive, but in recent years the transitivity of the causal relation has become controversial. (See Sartorio (2006) for an argument against the transitivity of causation, and Kvart (2001), Hall (2000), and Hitchcock (2001) for further discussion and debate.)

Many other problems have surfaced since Lewis's defence of the simple account, especially problems with varieties of pre-emption and overdetermination. As it turns out, the problems of pre-emption and overdetermination mean that the simple account is unworkable, and that a successful counterfactual analysis will need to do more than take the ancestral of dependence. Some of these issues are discussed in detail below, and the topic is dealt with at length in Hall and Paul (forthcoming). The many and varied problems with simple accounts have led to a proliferation of more complex counterfactual treatments of causation, for example, (p. 160) Ganeri, Noordhof, and Ramachandran (1996), Hitchcock (2008), Lewis (1986*c*; 2004*a*), McDermott (1995), Menzies (2004), Paul (2000), Schaffer (2005), Woodward (2003), and Yablo (2004).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

1. Methodology

Counterfactual analyses have received a good deal of attention in recent years, resulting in a host of counterexamples and objections to the simple analysis and its descendants. The counterexamples are often complex and can seem baroque to the outsider (indeed, even to the insider), and it may be tempting to dismiss them as irrelevant or uninteresting. But while we may be able to ignore some counterexamples because the intuitions they evoke are unclear or misguided, the importance of investigating the causal relation via investigating counterexamples should not be underestimated. Consideration of specific cases is extremely important, for such cases can discover a previously unrecognized general issue for counterfactual analyses, and often give us insight into the nature of the causal relation itself, independently of any particular analysis of it, counterfactual or otherwise.

In particular, the close investigation of counterexamples involving cases of pre-emption (e.g. where *C* causes *E* and *D* does not, but if *C* had not occurred, *D* would have caused *E*) and overdetermination (e.g. *C* and *D* cause *E* and neither *C* nor *D* causes the other, and if *C* had not occurred, *D* would still have caused *E*, and if *D* had not occurred, *C* would still have caused *E*) can give insight into many of the deep and delicate issues surrounding an account of causation. The need for clearer and deeper accounts of the transitivity of causation, what a causal process is, what it is to 'interrupt' a causal process, the role of omissions in causation, the role of the intrinsic character of the causal relation, and the connection between the intrinsicality of the causal relation and the character of the counterfactual dependence relation, was only recognized when counterexamples involving various sorts of pre-emption or overdetermination were discovered.

Apart from the need to investigate central counterexamples thoroughly, there are other methodological issues that affect investigations of causation and by extension any counterfactual treatment of causation. Recent work on causation has not been clear enough about the nature of the project or the problem that is to be solved. Is the project one of pure conceptual analysis, so that the problem is to develop an adequate analysis of a concept of causation? If so, is the concept of causation to be analysed a folk concept, a philosophical concept, a scientific concept, or something in between?

(p. 161)

There are more variations that need to be sorted: conceptual analyses could be descriptive, focusing on describing or elucidating our actual concepts, or prescriptive, focusing on constructing an improved version of a concept of causation. Prescriptive treatments can combine an analysis of our ordinary concept with other desiderata to construct a philosophical concept of causation, draw on developments in science in order to challenge certain assumptions implicit or explicit in our ordinary epistemic stance towards the world, and learn from psychology to help us identify features of our causal concept. Ideally, a prescriptive account will combine all these techniques in order to eliminate internal contradictions and construct a clean, clear, and sensible concept of the causal rela-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

tion that refines our unruly pretheoretical notions. Recent work in experimental philosophy draws on psychology for information about actual concepts: such work is obviously relevant to descriptive conceptual analyses, and may be relevant to prescriptive analyses as well (e.g. Hitchcock forthcoming, Hitchcock and Knobe forthcoming).

If the project is not conceptual analysis, is it ontological—a study of the ontology of the causation relation? Such a study could take the causal relation to be fundamental, or could take it to reduce to more fundamental entities. Further questions arise when we try to determine how we move from concepts to ontology (does a conceptual analysis give us an ontological analysis?) and what the role of our concept of causation is in determining the ontology of the relation. The best theories will probably mix some conceptual analysis with ontological analysis, if only to help clarify the subject under discussion. But what sort of a mix should we aim for?

Obviously, many people are interested in understanding how our concept of causation fits with the world, and to this extent we need to be clear on certain conceptual issues regarding causation even if our primary goal is ontological reduction. But the conceptual analysis of causation might also fit hand in hand with an examination of its ontology. Advocates who prefer to use conceptual analysis as a tool to carve out an ontological niche, Canberra-plan style (e.g. as in Jackson 1998), will defend a highly specialized account of conceptual analysis, usually one that is supposed to help us discover what is fundamental in the world via discovering what our words refer to. Such a plan for causation might proceed by analysing our concept of causation to define what the functional role of causation should be, and then taking the causal relation to be whatever relation (if any) in the world fills the role. One might also hold that conceptual analysis and ontological reduction are separate projects but necessarily related, such that success or failure in one sort of approach may translate into success or failure in the other.[2]

(p. 162)

A related approach might be to argue, following the structure of debates in the philosophy of consciousness, that conceptual analysis and related theses about conceivability and possibility can help us identify relevant metaphysically possible cases of causation.[3] One could even admit a role for conceptual analysis yet reject Canberra-plan approaches and other views with explicit requirements for how conceptual analyses are to be constructed and used. Such a rejection could still make room for some sort of conceptual analysis (which allows for attention to our philosophical intuitions about causation) to play an important role in helping us to distinguish between cases of causation and cases of non-causation. Conceptual analysis could then help to identify relevant features of the world needed to guide the development of a theory of causation and to target the subjects of the ontological reduction (see Hall 2006 for discussion of this idea).

Other positions about the interplay (or lack thereof) between conceptual analysis and ontological analysis can be staked out. One might wish to develop an analysis of a concept of causation while taking the ontology of any causal relation to be irreducible to any more fundamental relation. The goal of such an analysis would be to understand our concept or

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

concepts of causation, independently of what the metaphysics of the relation (should such a relation even exist) would involve. At the other end of the spectrum, one might not be interested in conceptual analysis at all, seeking answers about causation from fundamental physics or other empirical sources. Some of those who reject conceptual analyses but who also refuse to make philosophy into a handmaiden to natural science might draw on robust metaphysical theories about the structure of reality and the nature of possibility as well as empirical sources to develop an ontology of causation.

Moving on, another methodological issue involves the role of normative or pragmatic factors in one's philosophical approach. A counterfactual-based approach could rely on pragmatics to construct an analysis of causation (for a very nice sample of such work, see Hitchcock 2008). Such an approach involves adopting a measure of pragmatism such that subjective or normative elements are built into the account. Pragmatic accounts have much to contribute to the literature on causation. One may rely on pragmatics to determine that a series of events or causal processes are relevant or irrelevant (or whether causal pathways are ‘main’ pathways or ‘alternative’ pathways) with respect to our causal judgements, to decide whether an effect should be included in the representation of the causal structure of a case, to establish the values of event variables to be assigned in causal models, or to determine truth values for certain counterfactuals.

(p. 163)

However, these accounts face an objection: isn't such an account running together causal explanation with causation?[4] It seems right to say that causal explanation is mind- and description-dependent. It seems wrong to say that *causation* is mind- and description-dependent. Those who endorse a pragmatic account of causation will have to bite the bullet and deny this judgement. They may be willing to do so, given that pragmatic accounts will usually have a much easier time handling problems with pre-emption, overdetermination, and other serious difficulties for non-pragmatic reductive analyses. The ends may justify the means. But it is important to see that the metaphysician or philosopher of science who gives a pragmatic account of the causal relation or of the concept of causation is giving up on the hardest part of the puzzle: she is giving up on the analysis of the causal relation as an objective, that is, a description- and norm-independent, relation in the world or of the causal concept as a description- and norm-independent concept of an objective relation. For many philosophers of causation, giving an account of causation or of our concept of causation requires the presumption that the relation and our concept of it are objective.

The benefit of giving up on the hardest part is that the defender of the pragmatic account can usually solve the worst of the knotty problems raised by the possibility of pre-emption, overdetermination, causation by omission, and the like, or at least make sense of such cases in an intuitively plausible way. The cost is that our natural presumption that we have an objective concept of causation and that causal relations between particular events are objective relations, must be given up. Those who prefer to avoid infecting causation with pragmatics (keeping the pragmatics for, e.g., accounts of *explanations* given

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

in causal terms as opposed to accounts of *causation*) must continue to grapple with these exceedingly difficult counterexamples.

As I've indicated, then, the most ambitious accounts of causation are reductive. Reductive ontologies of the causal relation hold that facts about what causes what are fixed, somehow, by non-causal facts about what happens together with the facts about the fundamental (also non-causal) laws. Reductive conceptual analyses analyse our concept of causation in terms of other, more fundamental non-causal concepts. A reductive counterfactual analysis of the causal relation is an ontological analysis of how the causal relation reduces to ontologically more fundamental dependence relations, and a reductive counterfactual analysis of the concept of causation is an analysis of the concept in terms of (supposedly more fundamental) concepts of counterfactual dependence.

(p. 164)

An analysis of the ontology of the causal relation could aim for a purely local reduction: to reduce the this-worldly causal relation to more fundamental this-worldly relations (e.g. Fair 1979; Dowe 2000), or might aim for a reduction that encompasses the actual world and a closely limited set of worlds very much like our own. A more ambitious reduction would be a general treatment of causation in terms of (reductive) supervenience: such an account would specify the supervenience base for the causal relation such that any possible world that has the supervenience base has causation.

Many metaphysicians find reductionism very appealing, although some explicitly reject it (e.g. Carroll (Ch. 13 below); Tooley 1987; 1990; and Woodward 2003). More recently, counterfactual-based accounts that partly reject conceptual reductionism have received a good deal of attention, particularly causal modelling accounts such as Pearl (2000), Halpern and Pearl (2005), and Hitchcock (2001). Causal modelling relies on antecedent notions of what caused what, or on how to assign values to variables in the model in order to represent causal patterns and causal manipulability. Again, to understand the logical space of theories of causation, it is important to see that these non-reductive conceptual analyses of causation engage in a different sort of—admittedly important—project from that engaged in by the metaphysician who is attempting to give a non-pragmatic, reductive analysis of causal concepts and (especially) from the metaphysician who is attempting to give a non-pragmatic, reductive analysis of causal relations.

The methodological precepts of reductive views need to be clarified before one can determine whether a particular reductive project need—or need not—be reductive across the board. For example, if one's methodology requires everything in the world to supervene on patterns of contingently related instantiations of properties, then one's reductive theory of causation should include a reductive theory of laws of nature and an appropriately reductive account of modality. But a theory of causation can still be reductive without such broad reductionist commitments, for example, it could include a non-Humean account of laws or be primitivist about modality, as long as there were no stealthy assumptions about causation buried in these associated views. For example, any reductive ontological analysis of the causal relation will require the laws of nature that serve as part of

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

the supervenience base for causation in a world to be specifiable in non-causal terms, even if such laws are treated in non-Humean terms. Laws making explicitly causal claims such as '*x*'s having *P* causes *y*'s having *Q*' (e.g. as we see in Tooley 1990: 226) cannot be fundamental in a reductive treatment of causation: the acceptable law must be stated as 'if *x* has *P*, *y* has *Q*'.

No matter what method one applies to the analysis of causation, one must be clear about the role of thought experiments, intuitions, and the relevance of various conceptual and metaphysical possibilities. The rich history of conceptual analysis makes the role of such philosophical tools reasonably clear, even if issues such as the relation between conceivability and possibility remain controversial. (p. 165) But the difference between the goals and presuppositions of conceptual analysis and those of unabashed ontological analysis mean that the precise role of thought experiments, intuitions, and counterexamples could also be quite different in determining the success or failure of an account of causation. In other words, exactly what sort of project one is undertaking may influence one's approach to dealing with challenges to various proposed analyses, especially with respect to how a proposed counterexample needs to be addressed (this point is often lost among the thicket of the various analyses and objections proposed). For example, if an account of causation only aims to give an account of the causal relation in the actual world (and worlds with the same physical laws), then counterexamples involving bizarre laws of magic at distant possible worlds are irrelevant. Related issues crop up when deciding whether one is tackling the hard problem of causation or whether one is content to accept non-reductive or pragmatic elements as part of one's account. Philosophers working on ontological analyses (or analyses involving a mix of ontological and conceptual analysis) have not been precise enough about how the widespread rejection of pure conceptual analysis requires a change of methodological perspective.

A related problem for those working on the ontology of the causal relation concerns the breadth of their focus on ontology: are they content to rely on results derived from conceptual analyses for associated topics, such as the ontology of modality or of laws? If they reject, in principle, the project of constructing conceptual analyses as a tool for guiding reductions, then they should re-examine their commitments to related topics which may have been influenced by considerations motivated by conceptual analysis. For example, one might wish to revisit the semantics of counterfactuals or the theory of laws one endorses, since, for example, accepting more primitivism in one's accounts of what laws are or in determining the direction of counterfactual dependence can simplify or change the task for the ontologist of causation. If, instead, one's view is that conceptual analyses are appropriate guides for some reductions but not others, a consistent accounting of which topics are suitable for conceptual analysis—and why—is called for. The interaction between all of these methodological issues and accounts of causation (counterfactual or otherwise) is an area where further research is desperately needed.[5]

In what follows, I will focus on reductive, objective counterfactual analyses of causation and will try to be clear about the way that various methodological presuppositions infect our assessments of the strategies and problems for counterfactual analyses. I will distin-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

guish conceptual and ontological variants of these analyses of causation only when it is absolutely necessary.

(p. 166) 2. Motivation

Why attempt to develop a reductive conceptual analysis of causation or a reductive ontological analysis of causation? A general theoretical motivation for a reductive analysis of causation is that such an analysis would be deeply related to many other central philosophical topics, and would serve as a tool for philosophers, scientists, and others to use, the better to understand analyses of laws of nature, events, properties, practical reasoning, objects, probability, determinism, perception, mental causation, the existence of God, reference to and knowledge of the external world, agency, free will, moral responsibility, and legal responsibility.

Why attempt to develop a *counterfactual* reductive analysis of causation? One reason might be that there is no obviously superior alternative available. But a better motivation for adopting a counterfactual analysis starts with the intuition that cases of counterfactual dependence are cases of causal dependence. There seem to be systematic connections between our judgements that *E*s depend counterfactually on *C*s and our judgements that *E*s are caused by *C*s, and many theorists hold that the presence of counterfactual dependence (or its ancestral) is clearly sufficient for token causation. This is a connection between counterfactual dependence and the notion of influence and manipulability: if an event depends on another event, it is influenced or can be manipulated by that event. In everyday life as well as in the empirical and social sciences, causes are identified by the determination of manipulation: *C*s are causes of *E*s if changing *C*s changes the *E*s, that is, if we can manipulate *E*s by manipulating *C*s. In this way, experimental settings are designed to test for the presence of causation by testing for the presence of counterfactual dependence. (For discussion, see e.g. Winship and Morgan 1999.) The idea here is that the counterfactual dependence of *E* on *C* points to an underlying causal mechanism operating between *C* and *E*, and so detecting dependence is our best way reliably to infer the existence of the causal mechanism. (It is worth noting that Pearl 2000, Woodward 2003, and others have developed causally non-reductive interventionist accounts of causal modelling based on this idea.)

Relatedly, connecting causation with counterfactual dependence permits a 'black box' strategy that seems to be absolutely essential for our epistemic access to causation. If all we know about *E* is that it counterfactually depends on *C*, this is enough to infer that *E* is caused by *C*, even if we don't know anything else about the relation between *C* and *E*.[6] This is important for our everyday navigation of the world. If we had to know, for example, whether there was a process or a transfer of energy in order to know that there was causation, it would be much harder to recognize the causal relation. Since we seem to be able to determine the presence of (p. 167) causation just on the basis of dependence or manipulability, this gives us a strong prima facie case for the idea that mere counterfactu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

al dependence is sufficient for causation, and perhaps even for the idea that the causal relation is reducible to a (suitably qualified kind of) counterfactual relation.

A counterfactual analysis also has a reasonable amount of flexibility in responding to test cases, which is essential in order to treat crucial examples adequately. Lewis (1973*b*) argues that the counterfactual analysis, when combined with a sophisticated approach towards the semantics of possible worlds, correctly solves problems with common causes. Common cause cases are where *C* causes *B* and also causes *D* which, slightly later, causes *E*. *E* occurs later than *B*.

Consider the 'neuron' diagram of causation in Fig. 8.1. Capital letters name events. A filled circle represents the occurrence of an event, an empty circle represents the absence of an event, a line with a triangular head represents a causing of an event (a line with a circular head, not shown, represents the inhibiting of or the prevention of an event). Temporal progression from earlier to later is represented by reading from left to right.

One might think that *E* counterfactually depends on *B*, since had *B* not occurred, this would mean that *C* had not occurred, hence that *E* would not have occurred either. Not so, says Lewis: 'had *B* not occurred, *C* would not have occurred' is a *backtracking* counterfactual, counterfactuals that involve subjunctive conditionals where the time of the event described in the consequent precedes the time of the event described in the antecedent. Backtrackers are barred from use in evaluations of counterfactual dependence in worlds like our own. (Worlds with very different laws from ours might allow backtrackers. But in worlds with laws like ours, many backtrackers imply a violation of our laws, or at least of our laws plus certain assumptions about initial conditions.) Thus, even if *B* had not occurred, *C* would have occurred anyway, and *E* would have occurred, so *E* does not depend on *B*. If appropriate restrictions on backtrackers can be upheld, the counterfactual analysis rightly tells us that *C* is a cause of *E* but that *B* is not a cause of *E*.

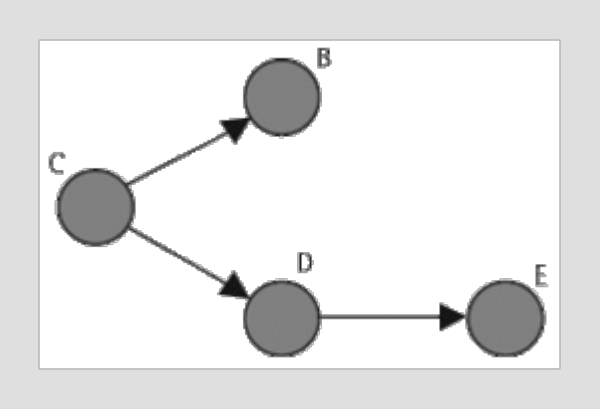

FIG. 8.1

(p. 168)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

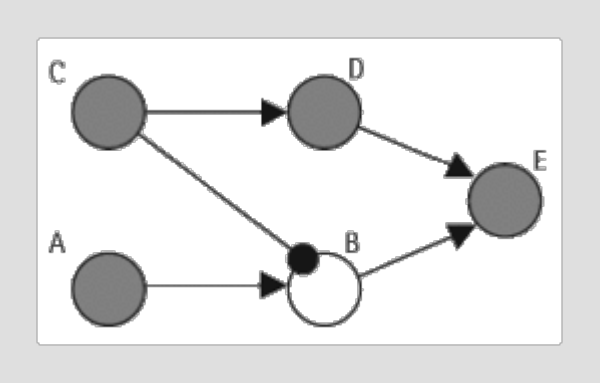

FIG. 8.2

A counterfactual analysis of causation can also handle certain sorts of cases of causal pre-emption; cases where events compete to cause an effect, but one or more of the competitors are pre-empted, in other words, the competitor does not succeed in causing the effect. Consider Fig. 8.2, which represents a standard case of causal pre-emption.

In Fig. 8.2, *A* and *C* are competing to cause *E*. *C* causes *E* while inhibiting *B*, thus pre-empting *A* (and *B*) from causing *E*. (The line that ends in a dark circle represents an inhibitory stimulus.) Assume that *E* would have occurred with exactly the same intrinsic characteristics if it had been caused by *A*. This case creates problems for simple counterfactual accounts, since *E* does not counterfactually depend on *C*.

Lewis's counterfactual analysis has a neat treatment of the case in Fig. 8.2. Since, on the assumption that causation is transitive, we take causal dependence to be the ancestral of counterfactual dependence, we see that *C* is a cause of *E* because there is stepwise counterfactual dependence between *C* and *E*.[7]

Another important motivation for a counterfactual analysis is that it can support causal claims in situations where there is no process, transfer of energy, or series of events between cause and effect. Such situations involve *negative* causation, that is, causation involving omissions or absences. For example, my failing to set my alarm causes me to miss my class, and my failure to hire a landscaper causes me to feel confused about what to do with my garden. The absence of a landscaper can cause my garden to run amok, and my inability to fix the garden can prevent my house from being sold. To the extent that these examples involve omissions or preventions of events, we can call them examples of negative causation.

(p. 169)

There are deep and puzzling issues surrounding causation by omission, since omissions, by definition, are not existents of any sort—they are *absences* of existences. Lewis (2004*b*) draws the striking conclusion that if omissions can be causes and effects, then there can be causation without a causal relation (because a relation requires relata, and at least one relatum goes missing if an omission is involved). For this reason, omissions

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

create especially difficult problems for conceptual analyses of causation that want to provide analyses of causation that do not separate off omissions as special cases.

Ways around some of the problems with omissions might involve translating claims about omissions into claims about certain regions of spacetime (although it can be difficult to develop a satisfactory account of which region or regions are being picked out, and to cover all cases, both possible and actual) or by defending fact causation, where the causal relation can take linguistic entities such as negative facts or sentences as relata. (For defences of fact causation, see McGrath (forthcoming); Mellor 1995; 2004.) Another issue involves the need to think in terms of types: it is usual to hold that an omission 'occurs' iff no event of a certain type occurs. But giving an adequate specification of such types is no easy matter. (One might think, in addition, that there is a normative component to the specification. See Beebee (2004) for critical discussion and McGrath (2005) for a defence of this idea.) These issues are pressing, serious problems for any analysis of causation, and much more research needs to be done to address the deep puzzles that causation by omission raises. But once we set these general problems with omissions for all analyses of causation aside, the important point that remains is that a counterfactual analysis of causation does a much better job of handling causation by omission than many of its rivals.

The accounts most threatened by causation by omission are those relying on transference, processes, or sufficiency based on actual events that instantiate properties of fundamental laws. Consider Fig. 8.3.

Take *C* to be the falling of a large boulder, and *A* to be Suzy's pushing Billy out of the way. *E* is the crushing of Billy under the boulder. The occurrence of *A* prevents the occurrence of *E*: Suzy saved Billy's life by preventing the crushing, that is, by causing an omission.

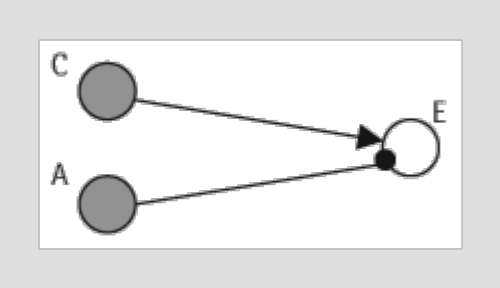

FIG. 8.3

This creates problems for any account of causation that requires there to be some sort of process or transference to link the causal relata or that cannot accommodate (p. 170) absences as causal relata (at least as relata in some sense). The problem particularly affects theories that reduce causation to processes or to sufficiency under fundamental laws, since unless there is a reconstruction of the causal sequence that involves an alternative process or a fundamental law, they must hold that no causation occurs.[8] (Dowe (2000: ch. 6) and Fair (1979) treat such cases in terms of different counterfactuals about process causation: it is unclear how they would handle Lewis's (2004*b*) example of the deadly void. For an alternative assessment of what is going on, see Beebee (2004).)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

Related cases such as the one in Fig. 8.4 involve what Ned Hall (2004) calls *double prevention*.

Such cases involve chains of (possible or actual) preventings (usually two in a row, hence 'double' prevention). A standard case involves Billy, who throws a rock at a bottle, and Suzy, who intends to prevent him by stopping his arm. As Suzy (*B*) reaches out to stop Billy's throw, Hillary reaches out (*C*) and prevents (*D*) Suzy from reaching towards Billy. Billy throws the rock (*E*). Hillary's grab prevents Suzy's block from preventing Billy's throw. In this case, Hillary's act is among the causes of Billy's throw, even though, as Hall points out, there is no process or fundamental regularity instantiated between Hillary's act and Billy's throw. Double prevention, once it is recognized, seems to be everywhere: guns firing, physiological processes, and everyday activities all involve double prevention.[9] It is simply unacceptable for an account of causation to hold that sequences involving double prevention are not causal.

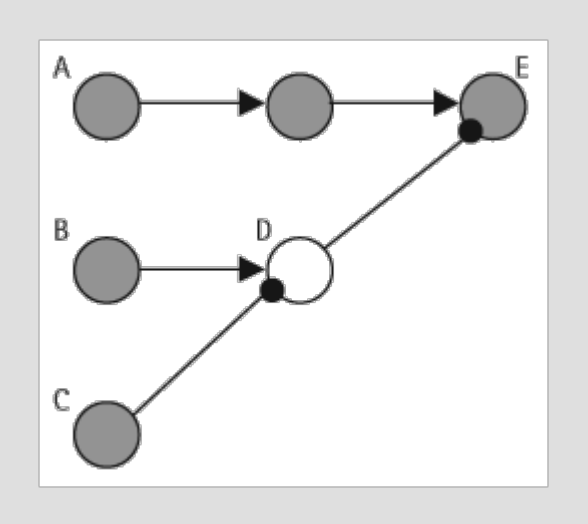

FIG. 8.4

(p. 171)

Because cases like those in Figs. 8.3 and 8.4 exhibit the requisite counterfactual dependence, the counterfactual theorist can hold, correctly, that such cases exhibit causation. If Suzy hadn't pushed Billy out of the way, he'd have been crushed. If Hillary hadn't grabbed Suzy's arm, Billy would not have thrown the rock. A counterfactual account makes such causation easy—all that is required is a certain sort of dependence—and thus can get the right answer when we have causation by omission. If, on the other hand, one requires that there be a transfer of energy or momentum between cause and effect for causation or denies that omissions can be causes or effects, cases of prevention and of double prevention don't count as cases of causation without a good deal more work.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

3. Problems

Although the elegance, relative simplicity, flexibility, and intuitive power of a counterfactual analysis are strong arguments in its favour, many problems have been raised in the literature. Some apply to counterfactual approaches generally, such as worries about circularity or the order of ontological dependence. (Which is more fundamental, causation or counterfactual dependence?) Others involve the way particular analyses deliver verdicts on cases that are at odds with our intuitions.

For reasons of space, I will consider only a few of the most central problems facing counterfactual analyses of causal concepts and the causal relation, and I'll look at this small sample in a fairly selective way. I'll discuss the general problem of circularity, and then consider the more specific and interrelated problems of pre-emption and overdetermination. As a result, I will set aside many problems and puzzles, including those involving indeterministic causation, distinctions between causation and causal explanation, trumping, the nature of the causal relata, further discussion of causation by omission, and transitivity. A close examination of the interrelated nature of pre-emption and overdetermination will help the reader to grasp some of the deepest and most central challenges for counterfactual analyses of causation: an extended treatment of many of the issues elided in this chapter can be found in Hall and Paul (forthcoming).

(p. 172)

3.1 Circularity

An account of causation based on counterfactual dependence requires a semantics of the counterfactual relation, that is, an interpretation of counterfactual claims that will determine the dependence relation that such claims rely upon. If the interpretation of counterfactual claims is determined by assessing causal claims, there seems to be a problem with circularity: one's account of the causal relation in terms of counterfactual dependence requires an account of counterfactual dependence in terms of causation. Perhaps such an account could still be informative: as long as we have enough conceptual access to causation to break into the circle, we can gain important insights by developing the tight connections between dependence and causation. (Woodward (2003) defends the view that a circular analysis is still informative.)

Lewis (1979) does seem to be employing causal notions to guide the development of his semantics of dependence, but such notions are used merely as rules of thumb to guide the development of the account of dependence in terms of qualitative similarities (of facts and laws) between worlds. The ultimate semantics is one that uses qualitative similarities as the ontological basis for evaluations of counterfactual dependence, and hence is not circular. (For details, see the discussion of the semantics of counterfactuals above.) Lewis's account is not without its problems, however, for as Elga (2001) argues, Lewis's account of the truthmakers for counterfactuals in our world is flawed. Lewis holds that certain asymmetries of overdetermination allow us to exclude backtracking counterfactu-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

als in a systematic way, which is important both for an adequate account of counterfactual claims in our world and for an adequate counterfactual analysis of causation for worlds like ours where there is no backwards causation. The problem is that, as Elga points out, Lewis's interpretation of how our world exhibits certain asymmetries of overdetermination conflicts with implications of our fundamental dynamical laws. As a result, depending on what sort of reduction or analysis one wishes to carry out, the semantics of counterfactuals and its attendant counterfactual analysis of causation needs to be patched up. In particular, advocates of a Lewis-style conceptual analysis of causation need to patch up the semantics for counterfactuals in a way that leaves it flexible enough to exclude backtracking conditionals for worlds like ours while still accommodating the possibility of having worlds in which there is backwards causation. The options and issues here are canvassed in Collins, Hall, and Paul (2004) and discussed in detail by Price and Weslake (Chapter 20 below).

(p. 173) 3.2 Pre-emption

Cases of pre-emption are cases where *C* causes *E*, but if *C* had not caused *E*, one or more back-up causes (merely potential causes) would have caused *E* instead. Such cases are cases of pre-emption (as opposed to cases of overdetermination—see sect. 3.3 below) because *C* is a cause of *E* while the back-up causes are merely *potential* causes. We have already considered a version of so-called *early pre-emption* with Fig. 8.2.

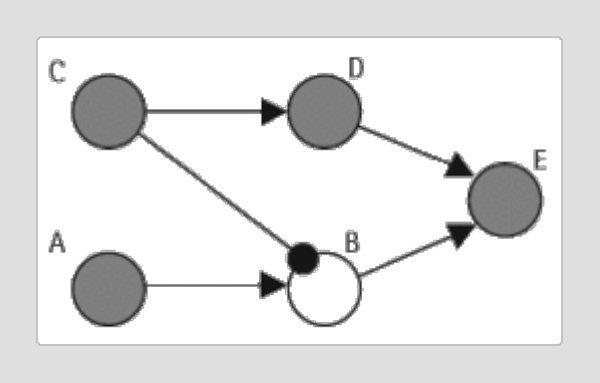

FIG. 8.2

The pre-emption is called ‘early’ because the interruption or modification of the back-up processes occurs before *E* occurs. A counterfactual analysis without bells and whistles tells us that *C* is a cause of *E* iff, if *C* had not occurred, *E* would not have occurred. But here is a case where if *C* had not occurred, *E* would have occurred anyway, albeit otherwise caused. Yet *C* is a cause of *E*.

The case brings out a deep problem for a counterfactual account: counterfactual dependence is sensitive to extrinsic factors such as extraneous events that function as possible causes or back-ups. In other words, the relation of counterfactual dependence between cause and effect can be affected by non-causal goings-on in the neighbourhood. The prob-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

lem is that we are inclined to judge that causation should not be sensitive to these sorts of extrinsic factors: intuitively, when *C* causes *E*, it does so whether or not there are other events around, assuming that these other events are not causally or otherwise necessarily connected to *C* or *E*. This is a way of saying that whether or not *C* causes *E* is independent of other entities not causally or otherwise connected to them (apart from the laws).

Yet we are trying to analyse causation in terms of counterfactual dependence. If so, how can such an analysis be correct? The most persuasive mitigating factor with regard to this problem is the competing intuition that counterfactual dependence seems to capture something deep and necessary about causation. If so, perhaps the (p. 174) answer is that an analysis of causation needs to have counterfactual dependence as a central part of it, even if the final analysis involves extra components designed to insulate (if such can be done) tests for dependence from extrinsic noise. While such an approach seems reasonable, counterfactual accounts have, so far, failed to make it work, making it unclear how to resolve these competing issues. (Though see Hall 2004 and Hall and Paul forthcoming for various suggestions about how to resolve the conflict.) This makes trouble no matter what sort of counterfactual analysis we are trying to develop, since a conceptual analysis of the concept of causation as a concept of counterfactual dependence would then seem to have mistargeted the relevant concept, and an ontological analysis would seem to have misidentified the reduction base.

Setting these deeper concerns aside, how do we address early pre-emption? In Fig. 8.2, the modification of the back-up process is an outright prevention: *C* interrupts the causal process between *A* and *E* by preventing *B*. However, cases of early pre-emption can involve other ways of modifying the causal process besides preventing an event. Perhaps, when *C* occurs, *C* modifies *B* in such a way that *B* still occurs, but would no longer be able to cause *E*. If so, then when *C* occurs, *C* is a cause of *E* and *A* and *B* are not. Yet, if *C* had not occurred, *E* would still have occurred, caused by *A* and *B*. See Hall and Paul (forthcoming: sect. 5) for discussion of just such a case.

One seemingly obvious quick fix won't do the job. You might think that whatever *A* and *B* cause, it isn't *E*, since *E* is something that is necessarily caused by *C*. This solves the puzzle by individuating *E* by its causes. But recall that we are trying to give an account of what causation *is*. We cannot individuate events by their causes and effects while reducing or analysing causation in terms of counterfactual relations between events, since without an independent account of which events occur, there is no way to determine which counterfactual relations exist!

Counterfactual theories of causation usually solve the problem of early pre-emption by relying on the transitivity of the causal relation. Instead of an analysis where *C* causes *E* iff, if *C* had not occurred, *E* would not have occurred, take causation to be the ancestral of the counterfactual relation, that is, take *C* to be a cause of *E* iff, if *C* is connected to *E* by a series of events $D_1, D_2, \ldots D_n$ such that if *C* had not occurred, D_1 would not have occurred, and if D_1 had not occurred, D_2 would not have occurred…and if D_n had not occurred, *E* would not have occurred. This modified counterfactual analysis, when applied

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

to the case in Fig. 8.2, takes *C* to be a cause of *E* because *E* depends on *D*, and *D* depends on *C*. So *C* is connected to *E* by a chain of counterfactual dependencies, even if *E* does not depend on *C* outright. (This solution only works if backtracking in relevant cases is prohibited: see Lewis (1973*b*) for discussion.)

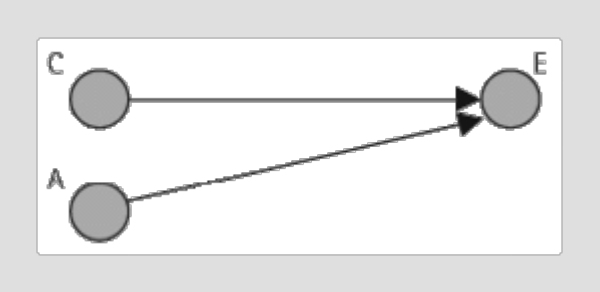

FIG. 8.5

Another sort of pre-emption that has received a lot of attention in the literature is *late pre-emption*. Late pre-emption is best described as pre-emption where *C* (p. 175) causes *E*, but pre-empted back-up processes are not interrupted until *E* occurs.[10] A textbook case of late pre-emption involves a pre-empted back-up causal process that is interrupted because the pre-empting cause brings about the effect before the back-up cause can (under the laws). Such a case is represented by Fig. 8.5.

Here, the pre-empting cause, *C*, causes *E* just before *A* would have caused *E* (this is represented by the arrow from *A* failing to extend all the way to *E* at the time *E* is caused).[11]

As with early pre-emption, the trouble is that *C* is a cause of *E*, but *E* does not depend on *C*: had *C* not occurred, *E* would still have occurred, since *A* would have caused it. But this trouble cannot be fixed by an appeal to transitivity, since whether or not there are events in the causal chain between *C* and *E*, there is no point at which *E* counterfactually depends on any such intermediate event (as it did on *D* in Fig. 8.2, above, in our sample case of early pre-emption). Thus we cannot take *C* to be a cause in virtue of being connected to *E* by a chain of dependencies.

Two quick fixes come to mind. First, one might try to individuate effects by their causes, so that *E* couldn't be caused by *A* after all. As I noted above, this approach cannot be used to solve pre-emption puzzles, since such puzzles arise within an attempt to develop a reductive analysis of causation. But there is a related move that might seem more promising: take events to be fragile, that is, to have extremely well-defined essences. I've already noted that in a textbook case of late pre-emption, *A* can't cause *E* when *C* causes *E*—if *A* had caused *E*, it would have caused *E* to occur a little bit later than it actually did. The fragility strategy exploits this fact by taking events to be temporally fragile, that is, taking events to be such that they could not have occurred any earlier or later than they actually did. If so, (p. 176) then in Fig. 8.5, *E* depends on *C*, since had *C* not occurred, *E* would not have occurred. Lombard (1986) and Coady (2004) develop fragility approaches.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

Many authors, most notably Lewis (e.g. 1986*b*; 1986*c*; 2004*a*), have been sceptical of the fragility solution, denying that events must have fragile temporal essences. If events are temporally fragile, then when that event occurs, it is impossible for it to have occurred earlier or later than it actually did. Note that simply accepting that temporally fragile events exist isn't enough to solve the problem, since it seems plausible, of course, that some temporally fragile events exist. The solution needed to solve the late pre-emption problem is not the thesis that there exist temporally fragile events, but rather that *every* event is fragile with respect to the time it is located at. Otherwise, it is easy to design counterexamples involving the pre-emption of one of the non-temporally fragile events.

There is, however, a more plausible alternative: instead of taking events to be temporally fragile, take causal counterfactuals to be fragile. (See Paul 1998*b*.) On this approach, we exploit the fact that if *C* had not caused *E*, it would have occurred later than it actually did, but without requiring a special metaphysics of *E*. Instead, we define causation such that (along with the usual caveats), had *C* not occurred, *E* would not have occurred *when it actually did*. This requirement correctly classifies *C* as a cause. However, the solution seems to be arbitrarily specific to times, and accordingly, Lewis (2000; 2004*a*) and Paul (2000) extend it to include counterfactuals that are sensitive to a wide range of characteristics, such that, had *C* not occurred just as it actually did, *E* would not have occurred just as it actually did. The final version of the view defended by Lewis is the thesis that, if whether, when, and how *C* occurs influences *to a suitable degree* whether, when, and how *E* occurs, *C* is a cause of *E*.

These proposals, while capturing a part of the content of our concept of causation, have been criticized. Although the influence of one event on another seems to be a significant marker for causation since it is the primary empirical tool for the determination of causation, it can lead one into error. Some object that using cases that purport to show that changing just any characteristic of an event is not enough to merit causal status. (Paul (2000) defends this result, given that causes and effects are property instances rather than entire events, while Lewis (2000; 2004*a*) denies it, adding the rider that the change in characteristics must be enough of a change to count as causation.) Others object that there are cases where *C* is not suitably influential in bringing about *E*, yet intuitively, *C* is a cause of *E*. A case of late pre-emption where the back-up process would have caused the effect to occur with the very same properties as *C* caused it to have, *modulo* reasonably plausible assumptions about what would have occurred if conditions were different, would be an example. These cases do seem to create problems for the influence theory. (For critical discussion of the influence theory, see Hall and Paul forthcoming; Schaffer 2001; Strevens 2003.)

(p. 177)

Two especially stubborn kinds of cases of late pre-emption—call them *esoteric* cases of late pre-emption—for any sort of counterfactual analysis that is supposed to be a conceptual analysis of causation involve (1) cases involving action at a distance and (2) versions of cases discovered by Goosens (1979) involving multiple or infinitely many pre-empted

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

alternatives. Lewis (1986*c*) considers instances of (1) but dismisses them as so far-fetched that his analysis need not address them. Given that Lewis is developing a conceptual analysis as opposed to an ontological reduction of causation, it would be helpful to have additional supporting arguments for this claim.

Even if we have a principled reason for setting aside cases involving action at a distance, esoteric cases of sort (2)—Goosens' cases—create trouble. To understand this sort of case, consider again the simple pre-emption diagram in Fig. 8.2. The causal chain started by C causes E, while the causal chain started by A is pre-empted. A Goosens-style case builds on this simple case by having many or infinitely many alternative back-up causal chains, each one being pre-empted just after the next. Instead of a single back-up, A, there are infinitely many (or as many as are needed, if time is discrete) backups $A_1, A_2, \ldots A_n$. The trouble here is that at every time during the causal process from C to E, there exists an uninterrupted back-up process. Such cases create problems for the transitivity solution to pre-emption since there is no point in the process from C to E where E depends on an intermediate event that depends on C. A third sort of esoteric late pre-emption is discussed by Paul (1998*a*) and Lewis (2004*a*) where there is no action at a distance, but where one cause pre-empts another without preventing events in the back-up process.

Strangely, Lewis (1986*c*) also dismisses Goosens-style cases as ‘too far-fetched’ and thus irrelevant to the success of his analysis. But it is unclear how such a verdict could be justified when one is pursuing a conceptual analysis: there seems to be no conceptual incoherence in the case, and so no reason to dismiss it as irrelevant. (After all, if cases involving magic and the like are acceptable, as Lewis (2004*a*) seems to think, then why not a Goosens case? Lewis (2004*a*) even accepts action-at-a-distance cases, effectively reversing his (1986*c*) view about cases of type (1).) We should take such cases seriously. A Goosens-style case where E does not depend on C because of the existence of back-up processes such that had any of those back-up processes caused E, E would have occurred exactly when and how it did occur, creates trouble for a broad range of conceptual analyses of causation in terms of counterfactuals.

3.3 Overdetermination

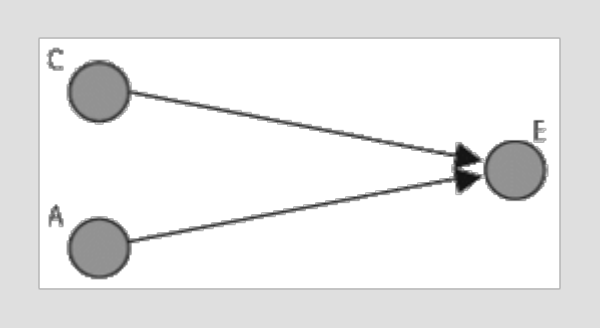

FIG. 8.6

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

Causal overdetermination in its standard form involves symmetrical causal contributions by multiple causal processes. Intuitively—but only very roughly—speaking, (p. 178) overdetermination can occur when more than one event, where each such event is part of a distinct, sufficient causal process, causes an effect. Fig. 8.6 gives a simple representation of a classic case of overdetermination.

In this case, *C* 's firing is sufficient to cause *E* 's firing in just the way it actually did. *A*'s firing is also sufficient to cause *E*'s firing in just the way it actually did. *C* causes *E* 's firing and *A* causes *E*'s firing. Although the intuitive idea of overdetermination seems simple enough, being precise about the deeper idea it involves, as well as getting the exact definition of causal overdetermination right, is extremely difficult.

To see why, start by considering the usual sort of case used to characterize overdetermination: two rocks, one thrown by Billy and one thrown by Suzy, hit a window at exactly the same time, shattering it. If this is a case of overdetermination, each rock-throwing alone causes the shattering. On a theory of causation where events are individuated robustly, that is, such that the very same shattering would have occurred whether there was one rock or two, this counts as a case of overdetermination.

But notice that if one counts the shattering by two rocks to be a numerically different event from the shattering by only one rock, and the world is such that if Billy hadn't thrown the properties of the shattering would have been different, and such that if Suzy hadn't thrown the properties of the shattering would have been different, then this is really a case of joint causation instead of overdetermination. (Billy and Suzy jointly cause the shattering, they don't overdetermine it.) For this reason, while the case of Billy and Suzy shattering the window can be treated as overdetermination, it is uninteresting to do so.

The interesting problems arise when we try to move beyond the sort of overdetermination we get when events are individuated robustly. For overdetermination with fine-grained events, where a difference in properties amounts to a difference in events, we need to describe an example such that that the shattering occurs *precisely* the same way, whether one rock shatters it or two. Call this *fine-grained overdetermination*. But surely, if the Suzy and Billy case is supposed to (p. 179) occur in (a deterministic version of) a world like our own, such a case is physically impossible. For example, each rock contributes a certain amount of force to the shattering of the window, and under the laws, the properties of the shattering will be affected by an increase or decrease in the force of the impact. So, assuming determinism, claiming that the shattering must be exactly the same whether there are two rocks or one requires a significant departure from (deterministic versions of) the laws of our world. (Paul 2007 calls this the problem of *additivity*.)[12]

Unless there are clean cases like those described in Fig. 8.6 above, where, for example, the firing of *E* is stipulated to occur when a certain activation threshold is met, and where *A* and *C* each contribute, in precisely the same way and at precisely the same time, sufficient activation energy for *E* to fire, and where neither *A* nor *C* is pre-empting the other nor jointly causing *E*, fine-grained overdetermination does not seem to be physically pos-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

sible (unless we can sort out a way to avoid the multiple subsumptions under the laws that additivity involves). For the example to work, don't think of *A* and *C* as each contributing a part of the energy needed to cause *E*'s firing—that way lies joint causation. Instead, by stipulation, *A* causes *E* in exactly the way it does in Fig. 8.6 even if *C* is absent, and *C* causes *E* in exactly the way it does in Fig. 8.6 even if *A* is absent.

In the end, such a clean case may not be physically possible. If so, this has implications for certain arguments about overdetermination and the actual world, especially when we consider debates about the possibility of mental or 'higher level' causation. On the assumption that token mental properties, events, or agents are not ontologically reducible to something more fundamentally physical, mental causation, agent causation, and other sorts of higher-level causation seem to involve fine-grained overdetermination. If so, and if such overdetermination is physically impossible, then these sorts of higher-level causation are physically impossible. (See Paul 2007 for further discussion.)

For the purposes of developing an analysis of causation, we can set this worry aside for now. Surely it is metaphysically possible to have a clean case of overdetermination, and, given a few adjustments to the laws, it is perhaps even physically almost-possible. If so, this sort of case needs to be addressed by any counterfactual account that takes itself to apply more broadly than to worlds with physical laws exactly like our own.

(p. 180)

The deeper issue underlying fine-grained overdetermination becomes clearer when we consider a clean case: how, exactly, can *A* and *C* each cause *E* if they are not causing it jointly? What work is being done by the much-needed stipulation that each cause brings about the effect just as it would if the other cause were absent? A way to express the worry is that it seems as though fine-grained overdetermination requires too much causation.

Note how differently we feel about the clarity of cases of fine-grained overdetermination versus that of cases of early and late pre-emption. In the pre-emption cases, the way each event makes (or doesn't make) its causal contribution is perfectly clear. But in fine-grained overdetermination cases, it just isn't clear how each cause is bringing about the effect all on its own, given that another cause is also bringing about the effect all on its own and the causation is not joint causation.

This is not an objection that assumes there is a certain amount of 'causal fluid' in the world that can bring about an effect and takes overdetermination to violate some sort of principle of the conservation of such fluid (see Sider 2003). Rather, it is the suspicion that there is a deep conceptual puzzle here about how the concept of overdetermination can fit with the concept of a sufficient cause as something that is entirely responsible for bringing about an effect. Given that *C* causes *E*, simply saying that *A* also causes *E* does not explain how *A* makes its causal contribution to the production of *E*. It merely reiterates that *A is* a cause of *E*. The worry is that fine-grained overdetermination seems to violate the natural intuition that—at least with cases that do not involve omissions—the com-

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

plete causal character of a causal chain is fixed solely by its intrinsic character plus the laws.[13]

Admittedly, given the variety of ways events can cause other events, intuitions about the role of intrinsicality in a causal analysis are controversial. But this controversy is mitigated to some extent once we see how, in a simple case like the one in Fig. 8.7, causal character is fixed by the laws and its intrinsic character.

FIG. 8.7

(p. 181)

In Fig. 8.7, how can it be the case that the causal facts about *E* can be changed without changing the intrinsic causal character of the causal chain between *C* and *E*? (The causal facts about *E* are changed because when we add *A* as a cause to get the case depicted in Fig. 8.6, *E* is overdetermined.) This simple case seems to show that intrinsicality matters.

Controversy about whether causal character is fixed solely by the laws and intrinsic character aside, the existence of the conceptual puzzle means that it is unclear, intuitively speaking, how fine-grained overdetermination works. Fans of counterfactual analyses try to exploit this lack of intuitive clarity, since counterfactual accounts have particular difficulties with cases of overdetermination. The main problem for counterfactual analyses derives from the sensitivity of counterfactual dependence to factors such as the presence of an overdetermining cause. *E* does not counterfactually depend on *C*, because *A* also causes it. Likewise, *E* does not counterfactually depend on *A*, because *C* also causes it. The defender of counterfactual analyses seems to be forced to fall back on one of two options: neither *C* or *A* caused *E*, for *E* is counterfactually dependent on neither of them, or the mereological sum of *C* and *A* caused *E*, for *E* is counterfactually dependent on this sum.

Holding that neither *C* nor *A* is a cause of *E* is unconvincing. *E* was caused, and *C* and *A* each seem to have caused it. How can it make sense to say that neither *C* nor *A* is a cause of *E* ? The latter option is more appealing: the *sum* of *C* and *A* caused *E*. But note that this won't get the counterfactual analyst as much as might seem. Taking the mereological sum of *C* and *A* as the cause does not mean that we are taking *C* as a cause and *A* as a cause: instead, the mereological sum *CA* is a cause while neither event alone is.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

This result is bizarre. How can the mereological sum of *A* and *C* be a cause of *E* while neither *A* nor *C* alone is a cause, joint or otherwise? It is unclear why the presence of *C* makes *A* unable to be a cause without being a member of the mereological sum of *A* and *C*, and likewise unclear why the presence of *A* makes *C* unable to be a cause without being a member of the mereological sum of *A* and *C*. Counterfactual accounts of causation violate the natural intuition that it seems right to say that if *E* is overdetermined then *A* is a cause of *E* and *C* is a cause of *E*, whether or not *A* and *C* compose a mereological sum.

For the counterfactual analyst who wishes to construct a reductive definition of overdetermination, the challenge is difficult. Further problems arise with the mereological sum treatment of overdetermination when we see that this is not enough to distinguish cases of overdetermination from joint causation, trumping (see Schaffer 2000*b*), or even early and late pre-emption. In order to avoid conflation with joint causation, one must add the requirement that whether, how, or when *E* occurs does not depend on *A* or on *C* (even while it does depend on their sum). One must also require that there be no interruption of the causal (p. 182) process between *A* and *E* and likewise between *C* and *E* (note that this is the only difference between the case represented by Fig. 8.5 and that represented by Fig. 8.6). Finally, one needs to add a requirement that *A* and *C* are distinct events, and that *A* does not depend on *C* and *C* does not depend on *A*. Moreover, all this must be done with non-circular definitions of 'interruption', 'causal process', and 'distinct'.[14] At this point, the prospects for a mereological sum treatment look grim, and it is unclear what sort of alternative treatments are available.

Further Reading

For early versions of counterfactual theories, see Lewis (1973*b*; 1986*c*), and Lyon (1967). For discussion of omissions, see Beebee (2004), McGrath (2005), Lewis (1986*c*), and Lewis (2004*b*). For discussion of problems with pre-emption, see Collins, Hall, and Paul (2004), Paul (1998*a*; 1998*b*), Lewis (1986*a*; 2004*b*), and Hitchcock (2007). For discussion of overdetermination and application to non-reductionism, see Paul (2007). For an in-depth cataloguing and assessment of these and other problems and issues involving counterfactual analyses of causation, see Hall and Paul (forthcoming).

References

BEEBEE, H. (2004). 'Causing and Nothingness', in Collins, Hall, and Paul (2004).

BENNETT, K. (2003). 'Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It'. *Noûs*, 37/3: 471–97.

—— (forthcoming) 'Two Causal-isms'.

CHALMERS, D., and **JACKSON**, F. (2001). 'Conceptual Analysis and Reductive Explanation', *Philosophical Review* 100: 315–61.

COADY, D. (2004). 'Preempting Preemption', in Collins, Hall, and Paul (2004).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

COLLINS, J., **HALL,** N., and **PAUL,** L. A. (2004). *Causation and Counterfactuals*. Cambridge, Mass.: MIT.

DOWE, P. (2000). *Physical Causation*. New York: Cambridge University Press.

ELGA, A. (2001). 'Statistical Mechanics and the Asymmetry of Counterfactual Dependence', *Philosophy of Science* 68/3 Suppl.: 313–24.

FAIR, D. (1979). 'Causation and the Flow of Energy', *Erkenntnis* 14: 219–50.

(p. 183) **GANERI**, J., **NOORDHOF,** P., and **RAMACHANDRAN,** M. (1996). 'Counterfactuals and Preemptive Causation', *Analysis* 56: 216–25.

GOOSENS, W. (1979). 'Causal Chains and Counterfactuals', *Journal of Philosophy* 76/9: 489–95.

HALL, N. (2000). 'Causation and the Price of Transitivity', *Journal of Philosophy* 97/4: 198–222.

—— (2004). 'Two Concepts of Causation', in Collins, Hall, and Paul (2004).

—— (2006). 'Philosophy of Causation: Blind Alleys Exposed; Promising Directions Highlighted', *Philosophy Compass* 1/1: 86–94.

—— and **PAUL**, L. A. (forthcoming) *Causation: A User's Guide*. Oxford: Oxford University Press.

HALPERN, J., and **PEARL**, J. (2005). 'Causes and Explanations: A Structural-Model Approach. Part I: Causes', *British Journal for the Philosophy of Science* 56/4: 843–87.

HITCHCOCK, C. (2001). 'The Intransitivity of Causation Revealed in Equations and Graphs', *Journal of Philosophy* 98: 273–99.

—— (2008).'Prevention, Preemption, and the Principle of Sufficient Reason', *Philosophical Review* 116/4: 495–532.

—— (forthcoming). 'Conceptual Analysis Naturalized'.

—— and Knobe, J. (forthcoming). 'Norms and Causation'.

JACKSON, F. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

KVART, I. (2001). 'The Counterfactual Analysis of Cause', *Synthese* 127/3: 389–427.

LEPORE, E., and **LOEWER**, B. (1987). 'Mind Matters', *Journal of Philosophy* 84/11: 630–42.

LEWIS, D. (1973*a*). *Counterfactuals*. Oxford: Blackwell.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

—— (1973*b*). ‘Causation’, *Journal of Philosophy* 70: 556–67.

—— (1979). ‘Counterfactual Dependence and Time's Arrow’, *Noûs* 13: 455–76.

—— (1986*a*). *Philosophical Papers II*. Oxford: Oxford University Press.

—— (1986*b*). ‘Events’, in Lewis (1986*a*).

—— (1986*c*). ‘Postscripts to “Causation” ’, in Lewis (1986a).

—— (2000). ‘Causation as Influence’, *Journal of Philosophy* 97/4: 182–97.

—— (2004*a*). ‘Causation as Influence’, extended version, in Collins, Hall, and Paul (2004).

—— (2004*b*). ‘Void and Object’, in Collins, Hall, and Paul (2004).

LOMBARD, L. B. (1986). *Events: A Metaphysical Study*. London: Routledge & Kegan Paul.

LYON, A. (1967). ‘Causality’, *British Journal for the Philosophy of Science*,18: 1–20.

MC DERMOTT, M. (1995). ‘Redundant Causation’, *British Journal for the Philosophy of Science* 46: 423–44.

MC GRATH, S. (2005). ‘Causation by Omission: A Dilemma’, *Philosophical Studies* 123: 125–48.

—— (forthcoming). ‘Uneventful Causation’.

MACKIE, J. L. (1965). ‘Causes and Conditions’, *American Philosophical Quarterly* 2/4: 245–64.

MELLOR, H. (1995). *The Facts of Causation*. London: Routledge.

—— (2004). ‘For Facts as Causes and Effects’, in Collins, Hall, and Paul (2004).

MENZIES, P. (2004). ‘Difference-Making in Context’, in Collins, Hall, and Paul (2004).

PAUL, L. A. (1998*a*). ‘Problems with Late Preemption’, *Analysis* 58/1: 48–53.

—— (1998*b*). ‘Keeping Track of the Time: Emending the Counterfactual Analysis of Causation’, *Analysis* 58/3: 191–8.

—— (2000). ‘Aspect Causation’, *Journal of Philosophy* 97/4: 297–328.

—— (2007). ‘Constitutive Overdetermination’, in J. Campbell, M. O'Rourke, and H. Silverstein (eds.), *Causation and Explanation*. Cambridge, Mass.: MIT, 265–90.

—— (forthcoming). ‘The Handmaiden's Tale’.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

(p. 184) **PEARL**, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

SARTORIO, C. (2006). 'On Causing Something to Happen in a Certain Way without Causing It to Happen', *Philosophical Studies* 129/1: 119–36.

SCHAFFER, J. (2000*a*). 'Causation by Disconnection', *Philosophy of Science* 67/2: 285–300.

—— (2000*b*). 'Trumping Preemption', *Journal of Philosophy* 97/4: 165–81.

—— (2001). 'Causation, Influence and Effluence', *Analysis* 61/269: 11–19.

—— (2005). 'Contrastive Causation', *Philosophical Review* 114: 297–328.

SIDER, T. (2003). 'What's So Bad about Overdetermination?', *Philosophy and Phenomenological Research* 67/3: 719–26.

STALNAKER, R. (1968). 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory*. Oxford: Blackwell.

STREVENS, M. (2003). 'Against Lewis's New Theory of Causation', *Pacific Philosophical Quarterly* 84/4: 398–412.

TOOLEY, M. (1987). *Causation: A Realist Approach*. Oxford: Oxford University Press.

—— (1990). 'Causation: Reductionism versus Realism', *Philosophy and Phenomenological Research* 50, Suppl.: 215–36.

WINSHIP, C., and **MORGAN**, S. L. (1999). 'The Estimation of Causal Effects From Observational Data', *Annual Review of Sociology* 25: 659–707.

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 Prototheory of Causation', in Collins, Hall, and Paul (2004).

Notes:

I am indebted to Peter Menzies and Jonathan Schaffer for spirited discussion, and to Daniel Silvermint for assisting with the bibliography.

(1) Stalnaker (1968) and Lewis (1973*a*) develop the accepted semantics for counterfactuals in terms of similarity of possible worlds to the actual world. See the introduction of Collins, Hall, and Paul (2004) for a more developed discussion of the evaluation of counterfactuals in the context of a counterfactual theory of causation.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

(2) Chalmers and Jackson (2001) hold that the success of an ontological reduction depends on the possibility of a certain sort of conceptual analysis.

(3) This idea is suggested by Karen Bennett in her forthcoming 'Two Causal-isms' and seems to be at work in discussions of trumping by Schaffer (2000*b*) and Lewis (2004*a*).

(4) Do not confuse the pragmatic approach with contextualist approaches (although the latter also face the objection that causation is conflated with causal explanation). Contextualist approaches are accounts where causal relations are indexed to contexts, much like velocities are indexed to frames of reference, and do not necessarily involve subjective or normative elements.

(5) I discuss some of these issues in more detail in 'The Handmaiden's Tale' (forthcoming).

(6) Lepore and Loewer (1987) argue, relatedly, that since dependence is sufficient for causation we can infer the presence of mental causation. See also Bennett (2003).

(7) Lewis (1973*b*) argues that a sufficiency account of causation, where C is a cause of E iff C is sufficient, under the laws, for E, cannot handle cases of pre-emption. Hall and Paul (forthcoming) argue that we should be cautious about the claim that the counterfactual analysis is an obvious winner with respect to the treatment of these sorts of cases.

(8) The sufficiency theorist may be able make use of derived laws, but must then explain the difference between derived laws and unlawful generalizations.

(9) Schaffer (2000*a*) describes how the interior mechanism of the firing of a gun, which might seem like a paradigm case of an uninterrupted causal process, actually involves a causal chain that includes double prevention.

(10) Terminological note: Lewis (1986*a*; 2000; 2004*a*) uses 'late preemption' differently from how I am using it: for Lewis, it refers only to cases where the causal chain is interrupted because the occurrence of the effect (caused by pre-empting cause *C*) prevents events in the back-up chain from occurring.

(11) Note that interpreting Figure 8.5 in this way involves a slight change in how the neuron diagram is being used—it represents the causal story at a time rather than over time.

(12) Indeterministic laws might not be violated, since it is consistent with such laws to hold that extremely improbable events are possible. So one might hold that it is extremely improbable yet physically possible that, in a case of pure overdetermination, an effect (such as the shattering) would occur just as it would have had it been caused by a single rock throw. However, the problem with additivity remains. In the case of indeterministic laws, the question we should ask is why the presence of an additional, overdetermining cause would not change the probabilities of the outcome. How could one lawfully claim that the probability of the shattering occurring (just as it actually did) is exactly the same

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020

whether there are two rocks or one? (Thanks are due to Jenann Ismael for raising the possibility of indeterminism in conversation.)

(13) The thesis about the dependence of the character of a causal chain on intrinsic features plus the laws is defended in Hall (2004).

(14) Even this may not be enough: in the end, in order to avoid esoteric counterexamples, the counterfactual analyst may be forced to rely on definitions involving intrinsicness, where, for example, the process between A and E and the process between C and E are required to exhibit certain intrinsic features or to intrinsically match certain causal processes. See Hall and Paul (forthcoming) for discussion.

L. A. Paul

L. A. Paul is Associate Professor of Philosophy at University of North Carolina at Chapel Hill. Her publications include ‘Aspect Causation’ in Journal of Philosophy 97 (2000); the co-edited volume (with John Collins and Ned Hall) Causation and Counterfactuals (MIT, 2004); and Causation: A User's Guide, co-authored with Ned Hall (forthcoming from Oxford University Press).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Yale University; date: 12 January 2020