Book Section:
Beebee, H orcid.org/0000-0001-7650-3622 (2022) The genesis of Lewis's counterfactual analysis of causation. In: Beebee, H and Fisher, ARJ, (eds.) Perspectives on the Philosophy of David K. Lewis. OUP . ISBN 9780192845443

Pre-publication version; do not cite. Published in: *Perspectives on the Philososphy of David K. Lewis,* eds. H. Beebee & A.R.J. Fisher (OUP 2022)

Chapter 11
The genesis of Lewis's counterfactual analysis of causation

Helen Beebee

1. Introduction

There aren't many philosophical claims that command close to universal assent, but over the last fifty years or so the claim that causation and counterfactual dependence are very closely connected has become one of them. Lewis's role in this—through his analysis of counterfactuals (1973a, 1979) and his development of counterfactual analyses of causation (1973b, 1986a, 2000, 2004)—cannot be overstated. The story of this chapter, however, is not the influence of Lewis's work on the topic of causation (see Bernstein 2019). It is rather the story of the historical context within which his game-changing paper 'Causation' (1973b) emerged, what the likely influences on his early thinking about causation were, and how that thinking developed.

In a letter written to Ken Kress in November 1972, Lewis describes the soon-to-be-written 'Causation' as his 'oldest piece of unfinished business' (Letter 21 in Beebee and Fisher 2020a: 37). And so it was: Lewis first proposed a counterfactual analysis of causation in a term paper, 'On Causality and Natural Laws' (1958, sometimes referred to in this chapter as 'Paper 1'), for his very first undergraduate philosophy class—Introduction to Philosophy, taught by Jerome Shaffer—in his second year at Swarthmore, a month or two one side or the other of his seventeenth birthday. This was followed by a second paper, 'Cause' (1959–60, Paper 2), written for his tutor Iris Murdoch during the 1959–60 academic year while he was a visiting undergraduate at Oxford studying PPE; and a third, 'Particular and General Causal Claims' (1961/2022, Paper 3), written when Lewis was back at Swarthmore.[1,2]

[1] Paper 1 was awarded an 'A' by Shaffer, with the comment 'Excellent paper'. There are three extant versions of the third paper, all very similar; one of them (probably the middle one) is dated 15 April 1961. The version I shall mostly quote from here—Paper 3 (Lewis 1961/2022)—is the most polished, and hence I am assuming that it is the final one. I also assume that this version was written in July 1961, but it is impossible to establish this with certainty; Frederique Janssen-Lauret and Fraser MacBride (2022) date it as 1965–6 on the grounds that it has 'Sent for job' handwritten by Lewis at the top, and it is unclear what job this might be unless an academic

All three papers are very short and cover roughly the same terrain. In a nutshell, they all introduce the idea that causation is a relation between events (or 'situations' (1959–60) or 'distinct situations' (1961/2022)) that is to be analysed in counterfactual terms; they say something about how to understand counterfactuals; and they discuss the epistemic and metaphysical relationship between causal claims that are 'particular' (the short circuit caused the fire) and those that are 'general' (short circuits cause fires). (In what follows I ignore general causal claims; I also follow early Lewis in using capital letters for the relata of particular causal claims.) As far as the evolution of Lewis's view from 1958 to 1961 is concerned, the three papers show a progressively decreasing concern with the niceties of various different and potentially non-equivalent causal locutions ('*A* caused *B*', *'A* was a cause of *B*', *'B* was caused by *A*', '*A* was the cause of *B*') and with general causation. And they show an increasing focus on developing and defending the idea that particular causation is a matter of counterfactual dependence, especially when it comes to redundant causation—a problem of which Lewis was apparently unaware until he read Shaffer's comments on his 1958 paper.

While the broader historical context within which these early papers were written is largely a matter of public record, how much of it Lewis was aware of is to a great extent a matter for speculation: his general penchant for the short bibliography and lack of detailed discussion of other philosophers' work was evident from the outset.[3] The bibliography for Paper 1—which unfortunately doesn't correspond to any actual mention of the authors in the main text—lists Russell's *Mysticism and Logic* (which includes 'On the Notion of Cause') and *Our Knowledge of the External World*; A.C. Ewing's 'A Defence of Causality' (1932–1933); and Herbert Feigl's 'Notes on Causality' (1953). Paper 2 contains no citations or bibliography, and the three versions of the 1961/2022 paper cite just two works: Hart and

---

job. But in any case this version is, as I say, very similar to the 15 April 1961 version, so even if it was written in 1965–6 it does not represent any remotely significant change in his views since 1961. In a letter to Michael Scriven (21 July 1961), Lewis says that he has submitted it—or possibly the 15 April version—for an essay contest at Swarthmore and that he 'hopes that it would be not too far from a publishable form' (Letter 2 in Beebee and Fisher 2020a: 4)—a forlorn hope, as it turned out, since it was subsequently rejected by both *Analysis* (in August 1961) and *The Journal of Philosophy* (in March 1962).

[2] All of the these papers are in the David Lewis Papers, C1520, Manuscripts Division, Department of Special Collections, Princeton University Library.

[3] In fact, Lewis engaged in a huge amount of detailed discussion of other philosophers' work; however he largely confined this discussion to his correspondence (see Beebee and Fisher 2020a and 2020b).

Honoré's *Causation in the Law* (1959) and R.N. Hanson's 'Causal Chains' (1955).[4] Nonetheless, there are other clues to be found along the way, as we shall see.

The chapter proceeds as follows. In §2, my basic aim is to show that Lewis's first attempts at a counterfactual analysis of causation constituted a bold and radical departure from the philosophical status quo of the mid-20th Century—though less of a departure from the legal status quo. In particular, Lewis's analysis completely abandoned the idea—dominant in philosophical contexts and increasingly influential in legal contexts—that cause and effect must be subsumable under some (exceptionless) law or regularity. In §3, I discuss some of Lewis's remarks on counterfactuals within the context of the broader mid-century debate on the topic, and begin to discuss the extent to which the seeds of his later (1973a) theory of counterfactuals were already present in his undergraduate work. §4, I continue the discussion of counterfactuals by considering Lewis's use of 'comparison cases' in evaluating counterfactuals. I explain how he uses comparison cases in his later undergraduate papers to try to overcome problems arising from redundant causation, and consider whether or not the early Lewis should be interpreted as offering a contrastivist account of causation. In §5, I briefly discuss his attitude towards the transitivity of causation. Finally, in §6, I discuss the period between Lewis's failure to get Paper 3 published in 1961 and the writing of 'Causation' (1973b), and note some important differences between his early and later work.

2. The break from sufficiency

It's fair to say that causation wasn't exactly a hot topic in analytic philosophy in the middle of the 20th Century. Hume had famously argued that inductive inference cannot be justified by appealing to causation; on the contrary, our causal judgements are a product of our inductive habits. Mid-20th-Century philosophers of science were very much interested in induction and confirmation but, taking their lead from Hume, took it for granted that worrying about causation in this connection would be of no help whatsoever. Thus, for example, causation gets barely a passing mention in Hempel's 'Studies in the Logic of Confirmation' (1945) or Popper's *The Logic of Scientific Discovery* (1959); and Goodman explicitly states at the start of 'The New Riddle of Induction' that 'Hume's dictum that there are no necessary connexions between matters of fact' has 'withstood all attacks' (1947: 59).

[4] Lewis briefly describes the broader influence that his year in Oxford had on him in a personal statement attached to a letter to the Dean of the Princeton Graduate School (Letter 3, 21 November 1961, in Beebee and Fisher 2020a: 6–10).

Causation, then, was largely either ignored—following the advice enshrined in the opening remark of Russell's 'On the Notion of Cause' that 'the word "cause" is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable' (1912–1913: 1)—or else (despite Russell's objections) viewed in a broadly Humean spirit, understood to be a matter of regularity and/or psychological projection of some sort.[5]

The 1953 paper of Feigl's that Lewis lists in the bibliography of Paper 1 falls into this latter category: 'The clarified (purified) concept of causation is defined in terms of *predictability according to a law*' (1953: 408). Ewing's (1932–1933) paper, however—which Lewis also lists—defies the orthodoxy. Ewing agrees with Russell that the regularity view of causation is hopelessly inadequate, but his main beef with it is, in essence, that things took a wrong turn with Hume: 'That causality is analogous to a logical connexion is confirmed by the fact that we can argue from the cause to the effect . . . I cannot possibly be entitled to argue from *A* to *B* unless I assume a connexion by which *A* really implies *B*' (1932–1933: 121). In other words, without real necessary connexions in nature of just the kind that Hume (allegedly) eschewed, induction is doomed.[6]

Within this intellectual context, to offer a counterfactual analysis of causation, as Lewis did in 1958, was in effect to break with the terms of the entire debate about causation engendered by the empiricist tradition of the first half of the 20th Century. After briefly raising some objections to the constant conjunction account of causation, Lewis suggests that we:

> define '*C* causes *E*' as 'Without *C*, no *E*, and with *C*, *E*'. This refers to single events only, not using classes of events at all. Thus the question 'Did the throwing of this stone cause the breaking of this window?' is equivalent to 'Did the stone get thrown, did the window break, and [i]f this stone had not been thrown, would the window have been broken?' (1958: 3).

---

[5] That Hume really did eschew necessary connexions is, however, a matter of controversy amongst Hume scholars—and indeed was roundly rejected by some Hume scholars in the first half of the 20th Century; see e.g. Smith 1905, 1941.

[6] It is also a matter of controversy amongst Hume scholars whether Hume himself thought that induction was doomed; see e.g. Garrett 1997: Ch.4 and Beebee 2006: Ch.3.

Thus conceived, particular causation—causation between particular events *C* and *E* (or 'situations' as Lewis calls them in Papers 2 and 3)—is completely divorced from regularity or lawful sufficiency or inductive inference between *C*-type events and *E*-type events. To use Lewis's example, that the signing of the Treaty of Versailles was a cause of the Second World War does not require—*pace* the regularity theory—that there is some regularity under which the treaty and the war can be subsumed; nor, *pace* Ewing, does it require that there is some quasi-logical relation between the two events, such that the first 'really implies' the second. Conceived as a matter of counterfactual dependence, causation is neither lawful sufficiency nor necessary connexion. That was quite a bold move in 1958. (And indeed in 1973.)

Or at least it was a bold move as far as philosophy was concerned. As far as legal practice and scholarship were concerned, the idea that causation is to be established on the basis of the 'but-for' or '*sine qua non*' test was utterly mainstream (though not entirely uncontested); it 'emerged unchallenged from the mists of time, entering the twentieth century as the only widely accepted judicial test of factual cause' (Rue 2003: 2684). The but-for test is explicitly counterfactual: as Joseph W. Bingham put it, 'a condition, force, or omission was a cause of a status if that identical status would not have occurred but for the contribution of the condition, force, or omission' (1909: 24). This counterfactual conception of (or, as it was usually conceived, test for) causation is routinely stated in legal textbooks and articles before and during the first half of the 20th Century.

There is no evidence, however, that Lewis was aware of any of this in 1958; indeed there is good reason to think that he was not. Discussion of the but-for test in legal textbooks invariably included the major problem with it, namely that of overdetermination—a problem that Paper 1 shows no knowledge of. Indeed, Shaffer's comment at the end of the essay presents a counterexample to Lewis's proposal in the form of an overdetermination case: 'Usually "cause" refers simply to sufficient conditions. Suppose a book rests on the table and a rope runs from it to the ceiling. Take away the table, and the book remains in place; take away the rope and the book stays in place. On your account, I take it, neither can be a cause, and yet surely either one is. Here each is sufficient, neither is necessary'. As we shall see later, Lewis attempts (not very successfully) to remedy this defect in his later undergraduate papers.

Lack of direct influence notwithstanding, there is a clear alignment between the interest in particular causation that inevitably arises in legal contexts and Lewis's own

interest in it—which differs markedly from the dominant interest of philosophers of the day in induction and confirmation. Lewis's paper starts:

> The problem of causation simply consists in finding a definition for the word 'causes' which fits our intuition, and is useful. We want it not to be the case that every event which occurs causes every other, and we want to have some empirical method of deciding, given two events, whether or not one caused the other. (1958: 1)

Deciding whether one particular event caused another is, of course, precisely what (*inter alia*) legal liability depends on. By contrast, the dominant view in philosophy at the time was, in effect, that the concept of causation wasn't *useful* at all—since such usefulness would consist in shoring up inductive inference and therefore playing a role in the justification of scientific theories, but empiricist scruples prevented (most) philosophers from conceiving causation in a way that would permit it to serve that purpose.

The lawful-sufficiency conception of causation that aligns with that dominant view was—as Russell had cautioned—completely unsuited to the task of deciding whether one particular event caused another, at least in cases where there are no obvious descriptions under which we might subsume those events in order to ascertain whether or not a regularity obtains between pairs of events so described (and according to Russell most cases in fact fall into this category). It was therefore unsuited to the task that was of interest both to Lewis and to the law.

The two very different traditions just described for thinking about causation came together to some extent with the publication of H.L.A. Hart and Tony Honoré's *Causation in the Law* in 1959, which Lewis had at least partly read by the time he wrote Paper 3; he had also attended Hart's lectures while studying at Oxford. In fact, the notion of 'causal sufficiency' was already well established in the law; overdetermination was generally conceived as a matter of the presence of two *sufficient* causes (as per Shaffer's comment on Paper 1), and this led to an increasing reliance, in US law at least, on the 'substantial factor' test for causation. The American Law Institute's Second Restatement of Torts (1965) says:

> If two forces are actively operating, one because of the actor's negligence, the other not because of any misconduct on his part, and each of itself is sufficient to bring about harm to another, the actor's negligence may be found to be a substantial factor in bringing it about [and is therefore a 'factual cause' of it] (American Law Institute 1965: §432(2))

The obvious problem with this statement is, of course, that any 'force' that arises from the actor will, on its own, *never* be sufficient for the harm: a fire requires the presence of oxygen, the damage to my car requires not just your speeding but the presence of my car as well, and so on. So the 'substantial factor' test came to be seen as requiring (at least) that the cause is a *necessary part* of a sufficient condition—a factor without which the sufficient condition would not have been sufficient.

Hart and Honoré's attempt to navigate the legal and metaphysical landscape involves bringing together the 'but-for' and 'substantial factor' accounts into a general account of what it is to be a 'causally relevant factor'. This delivers an account of what it is for a cause to be a necessary condition of an effect and hence, in effect, an account of what it is for an effect to counterfactually depend on a cause. Building on 'Mill's conception of the cause as a special member of a set of jointly sufficient conditions', Hart and Honoré list 'three different ways . . . in which a cause may be said to be a necessary condition of the consequence': (1) it is 'one of a set of conditions jointly sufficient for the production of the consequence' and 'it is necessary because it is required to complete the set';[7] or (2) 'it is a member of *every* set of sufficient conditions of a given type of occurrence'; or (3)—the *sine qua non* sense—it is 'necessary *on this particular occasion* for the occurrence of the event' (1959: 106–7). 'A condition which is necessary in any of the above three ways', Hart and Honoré say, is a 'causally relevant factor' (1959: 107).

Causes that are necessary conditions in sense (1) are typically so in sense (3) as well—but not in cases of overdetermination, where more than one jointly sufficient set of conditions is present. Overdetermining causes *C1* and *C2* will each be a necessary element in a sufficient set of conditions, and yet neither is a *sine qua non* of the effect. The problem with the legal *sine qua non* test of causation is thus that in capturing only one way in which a condition can be a causally relevant factor—namely (3)—it can yield 'misleading' answers (*ibid.*).

For Hart and Honoré, then, a particular effect *E* counterfactually depends on its cause *C*—that is, *C* is a *sine qua non* of *E*—if and only if *C* is a necessary element of a sufficient condition for *E,* where that sufficient condition—but no *other* sufficient condition—actually obtains. Despite the counterfactual form implicit in '*sine qua non*', the spirit of the regularity theory of causation is still alive and well in this proposal, since a sufficient condition for *E* is

[7] This, of course, is the basis of J.L. Mackie's 'INUS condition' account of causation (Mackie 1965).

to be understood as a set of event-types, one of whose members is a type under which *C* falls, that suffices for the occurrence of a type of event under which *E* falls.

Lewis, by contrast, never even seriously considers the possibility that *sine qua non*—that is, counterfactual dependence—might be understood in terms of regularity or sufficiency. As we saw above, he flat-out rejects the idea that we should use 'classes of events at all' in defining '*C* causes *E*', where '*C*' and '*E*' refer to 'single events' (1958: 3); and, as we shall see, the prototype counterfactual analysis he develops in that and the subsequent papers—unlike Hart and Honoré's way of understanding *sine qua non*—holds that line. Granted, generality—and indeed sufficiency—does enter the picture indirectly, since counterfactual dependence requires (to put it crudely) thinking away *C*, running the laws, and seeing whether or not *E* occurs (see §3 below). But this does not require that *C* is an event of a type that, on its own or in conjunction with other event-types, is sufficient for *E*-type events; it merely requires that the *absence* of the cause is lawfully sufficient for the *absence* of the effect.

Lewis's nascent counterfactual analysis of causation thus constituted a dramatic departure not only from the dominant view but from the whole way of framing the philosophical discussion of causation in the first half of the Twentieth Century.

3. Counterfactuals

By 1958 there was a lively debate about how counterfactuals should be understood, the issue having been addressed by, for example, Chisholm (1946, 1955), Goodman (1947), and Popper (1949).[8] One reason for the interest in counterfactuals at the time connects with the general interest in induction and confirmation: given the interest in the confirmation of laws, the question of how to account for the distinction between laws and merely accidentally true universal generalizations was a pressing one—and a tricky one, given the empiricist zeitgeist. Some philosophers—including Feigl in his 'Notes on Causality' (1953), which, as we saw earlier, Lewis lists in his 1958 bibliography—made the now-familiar suggestion that the distinction should be understood in counterfactual terms: laws 'support' counterfactuals while accidental generalizations do not. But in the absence of a viable analysis of counterfactuals it was unclear how this claim was to be made good.

The most promising proposal—pursued by both Chisholm and Goodman—was to think of the counterfactual conditional as entailment: $A > C$ is true iff *A*, along with some other true

[8] See Schneider 1953 for a discussion of the state of the debate at that time.

proposition or set of true propositions *S*, entails *C*. This proposal, however, is only viable if *S* includes one or more laws—otherwise laws will fail to support counterfactuals—but no accidental generalizations, otherwise those will erroneously support counterfactuals too. So this approach was a non-starter unless the law/accident distinction could be accounted for in some other way; and the prospects for doing so were not promising. As Chisholm put it:

> Can the relevant difference between law and non-law statements be described . . . without reference to counterfactuals, without use of modal terms such as 'causal necessity', 'necessary condition', 'physical possibility', and the like, and without use of metaphysical terms such as 'real connections between matters of fact'? I believe no one has shown that the relevant difference *can* be so described. (Chisholm 1955: 99)

Chisholm declines to offer such an account (1946: 302), while Goodman tentatively proposes a definition of lawlikeness: 'A sentence is lawlike if its acceptance does not depend upon the determination of any given instance' (1947: 23)—though unfortunately how we might define the circumstances under which this condition is met is 'a question I do not know how to answer' (1947: 27).

The other outstanding problem with the proposal was the 'cotenability' problem raised, but also not solved, by Goodman. Suppose we have a normal match in normal conditions, so that that the counterfactual (i) 'if the match had been struck, it would have lit' ($M > L$) is true but (ii) 'if the match had been struck, it would not have been dry' ($M > \sim D$) is false. Clearly, we need our set *S1* of statements that we conjoin with *M* to include *D* and exclude ~*L*, so that *S&M* will entail both *L* and *D*, delivering the truth of (i) and the falsity of (ii). But what prevents us from instead deploying set *S2*, which includes ~*L* and excludes *D,* which would make (i) come out false and (ii) true? After all ~*L* is true and compatible with *M* (and the laws), just as *D* is.

The only way of answering this question, as Goodman sees it (1947: 14–17), is to hold that *S* must be not merely consistent with but 'cotenable' with *M.* But the only way we can define cotenability in a way that gets the right result is by stipulating that it must not be the case that $M > \sim S$. Intuitively that gets us the right result: $M > \sim S1$ is false and $M > \sim S2$ is true. But now an obvious regress looms; as Goodman puts it, 'to establish any counterfactual, it seems that we first have to establish the truth of another'. Hence 'the problem of counterfactuals must remain unsolved' (1947: 16–17). And Goodman does 'not at present see any way of meeting' this 'difficulty' (1947: 17).

There is no evidence that Lewis was aware of any of this literature in Paper 1—not from his brief positive proposal, which I'll come to in a moment, nor from the other relevant passage, which is the paper's final paragraph:

> It will be seen that the above definition of causality depends completely on some other meaning of the counter-factual other than material implication, which would make all counter-factuals true. However, the concept of causality is not alone in this difficulty. I do not propose to discuss that problem here. (1958: 6) [9]

Paper 2 repeats more or less the same positive proposal; in Paper 3 the proposal is condensed to a single sentence and is accompanied by a similar caveat, though without mention of the material conditional. Perhaps by that point he had become acquainted with some of the literature and come to realize both that the point about material implication wasn't worth mentioning—that suggestion had been taken seriously by nobody, having been consigned to the dustbin at the start of both Chisholm's and Goodman's early papers (Chisholm 1946: 294–5; Goodman 1947: 4)—and that coming up with a viable analysis of counterfactuals was going to be rather trickier than he had hoped.

In any case, my interest here is in Lewis's positive proposal, of which the Paper 2 version is:

> The counter-factual ['without one, not the other'] is verified by prediction according to some sort of laws from the total situation initially of as large a part as necessary of the universe with the first situation extracted to the non-occurrence of the second situation. Just as for any prediction, laws of appropriate level are used; i.e. the highest possible amount of detail in which the calculation is still possible; with more detailed laws the certainty of the conclusion increases. For instance, to find out whether the signing of the Treaty of Versailles caused the Munich conference we pretend that the treaty was not signed and try to predict from what we know of social behavior what would have happened. We do not try to make our prediction from the laws of particle dynamics. The

[9] Modest evidence that he was not acquainted with the debate about counterfactuals comes from the fact that he hyphenates 'counter-factual'—not a practice followed by the main participants in that debate at the time (though admittedly not universally shunned either).

> calculation would be impossible, and the initial data would be impossible to collect. (1959–60: 2)

The basic recipe for the verification of counterfactuals that Lewis is offering here, then, is: take the state of the Universe at the time of the antecedent 'situation' *A*—or as much of it as is required—and 'extract' *A* from it. Then run the laws (of the 'appropriate level'). If it follows that the consequent situation *C* does not obtain, $\sim A > \sim C$ is verified.

While there is much more that could be said about this passage, I'll confine myself to a couple of observations. First, Lewis's focus on the *verification* of counterfactuals is of course in step with the post-positivist zeitgeist of the time. On the other hand—whether by accident or design[10]—he is leaving open the possibility that the truth conditions of counterfactuals come apart from their verification—that is, the possibility that, while we might *verify* counterfactuals via the predictive method just described, their truth conditions need not (*pace* Chisholm and Goodman) be conceived in terms of an entailment relation between the antecedent and some additional set of cotenable statements on the one hand and the consequent on the other. And that is precisely the option he takes in the possible-worlds account he later develops in *Counterfactuals* (1973a) and elsewhere: the truth of counterfactuals is a matter of what is true at the most similar antecedent-worlds, and not what is entailed by some set of statements or propositions.

Second, Lewis is in effect baking in (or better: half-baking in) a temporal asymmetry that was absent from earlier metalinguistic accounts. Indeed, what Goodman viewed as the main impediment to solving the cotenability problem was precisely one that only arises if temporal asymmetry is not baked in. Recall the problem: 'the match did not light' ($\sim L$) must fail to be cotenable with 'the match was struck' (*M*) or else 'if the match had been struck, it would have lit' ($M > L$) will come out false. But Goodman thought the only way to secure this result was to restrict cotenability to the set of true statements *S* such that it is not the case that if *M* were true, *S* would not be true—resulting in a vicious regress. Lewis's view gets around that problem by, in effect, stipulating that facts about the future relative to $\sim C$ should not be held fixed when evaluating the truth of $\sim C > \sim E$: when we excise *C* and run the laws, we hold fixed the 'total situation initially of as large a part as necessary of the universe with the first situation extracted'. The 1958 version is clearer on what he has in mind: Lewis there

[10] Perhaps more by accident than design; much later, Lewis does describe his 1961 self as a verificationist (Letter 523 to Robert C. Stalnaker, 26 February 1999, in Beebee and Fisher 2020b: 194).

stipulates, in effect, that we hold fixed 'the state of affairs *at the time of* the alleged cause' except that 'the alleged cause is stricken out' (1958: 5; my italics).

Temporal asymmetry is only half baked-in here, however, because, while Lewis is in effect demanding that we don't hold the future fixed, he *isn't* demanding that we *do* hold the past fixed—only the state of the Universe 'at the time of' the alleged cause. He hasn't, apparently, seen that this is problematic: if we strike out the cause and run the laws backwards in time rather than forwards, we will of course get true backtracking counterfactuals, which will in turn deliver spurious causation. Nonetheless, Lewis's basic recipe here bears an obvious affinity with the analysis of counterfactuals he later develops. In particular, the similarity metric of his 1979 (472; cf. 1973b, 566–7) turns a similarity-of-worlds version of the same trick when it comes to Goodman's problem case: once perfect match of matters of particular fact is lost, as it is at the time of *M* (the non-actual striking of the match), future similarity—and hence whether or not it lights—counts for nothing. And since the obtaining of *M and* something that would prevent it from lighting—such as the match's not being dry—requires two law-violations rather than just one (one to make *M* obtain and a second to make the match damp), the closest *M*-world will be one in which the match is dry and hence lights; so $M > L$ is true (and $M > \sim D$ is false) as required.[11]

In the discussion of counterfactuals so far, however, I have omitted one significant detail of the latter two of Lewis's early papers. It's time to factor that in.

4. Comparison cases and redundant causation

The significant detail just mentioned is that in Papers 2 and 3 Lewis is exercised by the question of what we put in its place when we counterfactually suppose away the cause. In Paper 2, he says:

> If the initial situation is extracted, what is put in its place? Often there is a plausible alternative; somebody is born or not. A flask is heated or not. A book is read by a man at a given time or he does something else, something which as far as our predictions go, we can neglect. But, strictly speaking, we should say that situation *A* (instead of *C*) caused situation *B*. (1959–60: 2)

[11] I am oversimplifying here, since later Lewis does not in fact 'bake in' temporal asymmetry by fiat. His 1979 similarity metric deliberately does not stipulate that the region of perfect match of particular fact must lie in the past rather than the future; rather, the required temporal asymmetry is the (alleged) upshot of what he calls the 'asymmetry of overdetermination' (see his 1979: 472–5).

In Paper 3, Lewis spends more time on what he is now calling 'comparison cases': what we put in its place when we 'imagine away *A* and replace it by situations (here called comparison cases) which could not fall under the description specifying *A*' (1961/2022: XX):

> What can be a comparison case? To know what would have followed, we commonly have to know more than that *A* was not the case. It may sometimes be that no situation without *A* would be followed by one with *B*; but we need make no such strict requirement in order to make causal claims. There are ways of telling what are and are not proper comparison cases. The sick man wasn't given the drug; but he was kept in bed, not made to run two miles . . . .There was no Treaty of Versailles; and neither was there another treaty of precisely the same wording but signed in a different place, nor was there an even harsher (in some contexts, milder) treaty. Sometimes we cannot tell what would be proper comparison cases. Indeed there is a locution to deal with this difficulty: 'That *A* (instead of *C*) caused situation *B*'.
>
> Usually we do not have to consider a multitude of comparison cases. Often there is only a single one, and it is quite plain what it must be. If we ask whether a change was caused by a drug, the comparison case is that the drug was not given, but everything else was the same. (There may be a question whether a placebo was given instead.) (1961/2022: XX)

There are two ways we might interpret Lewis's proposal when it comes to comparison cases, one of which aligns more closely with his later view on counterfactuals than the other.

The former interpretation takes the proposal to be, in effect, a mechanism for resolving the vagueness of (what he later calls) the similarity relation: we have a 'choice' of comparison cases just when there is no fact of the matter about which of (say) situation *C* and situation *D* is closer to actuality (where neither *C* nor *D,* but rather *A*, actually occurs). Often the indeterminacy doesn't matter, because the relevant causal claim ('*A* caused *B*') will come out true no matter which of the available choices we make. Thus, for example, suppose it's indeterminate whether, had he not finished reading the murder mystery this afternoon, the man would instead have gone for a walk or watched TV or listened to a concert on the radio. 'The man's finishing the book caused him to find out that the butler did it' will come out true whichever of these alternatives we choose as our comparison case, since in none of those scenarios would he have found out that the butler did it. So there is no need to make a choice

at all. In some cases indeterminacy does matter, however, because '*A* caused *B*' will come out true on one resolution of indeterminacy and false on the other. In such cases we can resolve the indeterminacy by making an explicit choice—that is, by specifying one comparison case rather than the other in our causal claim, so that '*A* (rather than *C*) caused *B*' is true but '*A* (rather than *D*) caused *B*' is false.

On this interpretation, the proposal falls within the same ballpark as Lewis's 1973a view of counterfactuals, where relative closeness of worlds goes by comparative overall similarity and there is often indeterminacy in such matters—an indeterminacy that can often (though not always) be resolved by context (see e.g. Lewis 1973a, 66–7).

On the other interpretation Lewis's proposal is a genuinely contrastivist one, in the sense that the existence of a choice between comparison cases is not limited to those cases where (again, to adopt Lewis's later terminology) it is indeterminate which situation—*C* or *D*, say—is closer to actuality. While the proposal, thus interpreted, is not merely that *causation* is a contrastive matter—it's rather that (what Lewis will later call) counterfactual dependence is itself contrastive—it is in the same ballpark as much more recent contrastivist accounts of causation developed by Cei Maslen (2004), Jonathan Schaffer (2005) and Robert Northcott (2008).

The basic idea of contrastivist accounts is that (a) causation is in fact a three-place rather than a two-place relation, as expressed by claims of the form '*c* rather than *c** caused *e*',[12] (b) the truth of such claims is determined by the truth of the counterfactual 'had *c** rather than *c* occurred, *e* would not have occurred', and (c) when it comes to an ordinary, two-place causal claim ('*c* caused *e*'), the claim is true just if, in the context of utterance, '*c* rather than *c** caused *e*' is true for the contextually salient *c**. In other words, context fixes which contrast case is salient, and that in turn fixes which counterfactual determines the truth of '*c* caused *e*'. Suppose, for example, that Jane steals a cake from the bakery and gets arrested. Thanks to the heavy security she would have been arrested whatever she'd stolen—but not, of course, if she'd bought the cake rather than stealing it. Thus the contrastivist account allows us to distinguish between 'Jane's *theft* of the cake caused her arrest' (true) and 'Jane's theft of the *cake* caused her arrest' (false): the stress in the first sentence selects Jane's buying the cake as the relevant contrast, whereas the stress in the second sentence selects her theft of some other item.

[12] Actually in Schaffer's and Northcott's versions it's a four-place relation: context fixes a salient effect-contrast (*e**) as well as a cause-contrast (*c**). I ignore that complication here.

The basic difference between the two interpretations amounts to whether or not the existence of a choice between comparison cases is constrained by how much of a departure from actuality is involved. On the former interpretation we have a choice between comparison cases when, and only when, it is indeterminate (absent a context to resolve the indeterminacy) which of two or more comparison cases constitutes the most modest departure; the explicit *mention* of comparison cases thus constitutes a way of forcing the resolution of indeterminacy one way rather than the other. The latter interpretation, which aligns Lewis's proposal with contrastivist accounts of causation, is much more permissive. On this view, the existence of more than one legitimate comparison case is not constrained by closeness: '*c* rather than *c** caused *e*' can perfectly well be true even if the closest *c**-world is considerably further away from actuality than many other worlds where *c* fails to occur. In the above example we might suppose that Jane is intent on stealing and had no intention of buying anything. In that case, it might be that all worlds where she buys the cake are considerably further away than many worlds in which she steals (say) a bun or a croissant, but cake-buying still constitutes a 'proper' comparison case. The truth of the causal claim depends not on what happens at the closest world where the putative cause *doesn't* happen, but on what happens at the closest world where the contextually salient contrast *does* happen—whether or not there happen to be determinately closer worlds where some other, non-salient comparison case occurs.

The immediate textual evidence, especially from the long passage from Paper 3 quoted above—and especially Lewis's remark that '[o]ften there is only a single [comparison case], and it is quite plain what it must be'—suggests the former interpretative option. And, as I'll now argue, Lewis's attempt to solve the pre-emption problem provides some additional support for that option.

As I noted in §2, Lewis seems to be unaware of the problem of causation without counterfactual dependence in 1958. Shaffer raises the overdetermination problem in his comments on that paper, and Lewis duly addresses it in Paper 2. He introduces it by noting that 'situations' can be conjunctive and disjunctive, and uses the example of Jones—who takes a cold bath and then goes for a walk in the rain, and subsequently catches flu. In the case where either the bath or the walk 'would have been enough by itself to cause the flu', he says:

> Then the bath did not cause his attack of the flu; neither did the walk. The bath and the walk together caused his attack of the flu [conjunctive situation]—but this is misleading,

> because it sounds like the first case in which they were both necessary. 'Either the bath or the walk caused the flu' [disjunctive situation] sounds as though one or the other did, and we could find out which. The best that standard usage gives us is 'The bath was *enough* to cause the flu, and so was the walk'. Or, 'Had he not taken the bath, the walk would have caused it, and *vice versa*'. (1959–60: 3)

By 1961, Lewis has clearly discovered a further species of (what he will later come to call) redundant causation, namely pre-emption. This time he distinguishes between (what he will later call) overdetermination and pre-emption, and gives an example in which two drugs (call them Drug *A* and Drug *C*) are administered ('situations' *A* and *C*) and the patient recovers (*B*). In the overdetermination version of the case, this time we see more or less the same sanguine attitude that comes out in his 1973 paper, where (what he there calls 'symmetric') overdetermination is relegated to a footnote, such cases being 'useless as test cases because I lack firm naïve opinions about them' (1973b: 567*n.*). In Paper 3, he says: 'We wouldn't really want to say that the first drug caused him to get well; but neither would we want to say that it was false that it did' (1961/2022: X). And he leaves matters there.

Lewis takes pre-emption more seriously, however, and his 1961 solution to the problem is to claim that '[o]ur answer to the question whether *A* caused *B* depends on a somewhat open choice of comparison cases' (*ibid.*). Thus if we know that administering Drug *A* deactivates Drug *C* (so that we know that this isn't a case of overdetermination), it is legitimate both to affirm and—by making a different choice—deny that *A* caused *B*. For we might choose as our comparison case one where, in the absence of *A*, Drug *C* is not deactivated—in which case '*A* caused *B*' comes out false because *B* would still have happened in the absence of *A*—or we might instead choose the comparison case where Drug *C is* somehow still deactivated (though of course not by Drug *A*), in which case '*A* caused *B*' comes out true.

This solution, such as it is, depends on the idea that we are ambivalent about what to say in pre-emption cases—and Lewis claims that we *are* ambivalent: 'it is legitimate to say that it was the first drug that caused the cure. It is *also* legitimate to deny it, appealing to the counterfactual, that the cure would have happened anyhow' (*ibid.*). One might wonder whether he is illicitly putting his theoretical cart before the intuitive horse here, but he does appear to be sincere: he expresses the same ambivalence in a letter to Michael Scriven,

written in February 1961 while he was likely to have been working on Paper 3.[13] He says: 'When it would have happened anyway, but in a different way, I may accept a causal claim in a coroner's report or a historical account . . . but am also prepared to reject it' (Letter 1 in Beebee and Fisher 2020a: 3).

But what exactly is the source of our alleged ambivalence? The two interpretations of the role of comparison cases in counterfactuals floated above give us two different answers. According to the first interpretation (again, adopting Lewis's later terminology) the ambivalence is due to the vagueness of the similarity relation: it is indeterminate whether the world (comparison case) where Drug *C* is somehow or other neutralised (call this w_1) is more of a departure from actuality than is the world where Drug *C* is not neutralised (call this w_2). According to the second interpretation, the ambivalence is due to failure to specify a context that selects one world rather than the other, much as we are (according to contrastivists) ambivalent about whether or not Jane's theft of the cake caused her arrest (yes if the contrast is her buying the cake; no if it's her stealing something else).

We might rule out the first interpretation on the grounds that it involves attributing to Lewis the highly implausible view that it's indeterminate whether w_1 or w_2 is more of a departure from reality. But in fact he raises this issue in the fourth of his Gavin David Young Lectures, written some time between 1968 and 1970, by which time he had 'up my sleeve an analysis of counterfactuals' (Lewis 1971/2022: XX). He discusses basically the same case, except this time involving poison and death rather than drugs and recovery: what is at issue is whether the closest possible world is one where the second poison—which is actually neutralised by the first poison—is still somehow neutralised (so that Oscar doesn't die) or instead one where it isn't neutralised (so that he does). He says:

> I might claim (not quite so implausibly) that it is allowable either to affirm or to deny that he would have died anyway, within the limits of vagueness of our concept of comparative similarity of worlds; and that it is correspondingly allowable either to deny or to affirm that my poisoning of his coffee caused his death. (Lewis 1971/2022: XX)

Granted, Lewis is conceding that this move is somewhat implausible, and he is clearly scratching around for a better solution to the pre-emption problem. (The other solution he floats on the same page is to claim that Oscar would still have died but would have died a

[13] Scriven was a visiting lecturer at Swarthmore from 1956 to 1960.

different death, so there is still counterfactual dependence between the administering of the first poison and the death.) Nonetheless, the above passage does provide some retrospective evidence for the first interpretation of his earlier view about 'comparison cases', namely that the 'choice' of comparison cases arises from indeterminacy when it comes to similarity with the actual world, and not— as per the second interpretation—from a contrastivist conception of counterfactual dependence.

One might object that the second interpretation provides a *better* solution to the pre-emption problem, since it doesn't rely on the dubious claim about indeterminacy just described; hence, by a principle of interpretative charity, we should ascribe that interpretation to Lewis. Well, granted, it doesn't rely on that dubious claim. Nonetheless, it's still a pretty poor solution to the problem. After all, 'Drug *A* (rather than no Drug *A* but with Drug *C* somehow neutralised) caused the death' seems somewhat unnatural, to put it mildly. 'Drug *A* (rather than no drug at all) caused the death' sounds better and gets us counterfactual dependence as required—but unfortunately 'Drug *C* (rather than no drug at all) caused the death' also comes out true if we're allowed to count the no-drug situation as a proper comparison case.

Perhaps in the end the right thing to say is simply that Lewis was just not sufficiently in command of his own material in 1961 for it to be appropriate to ascribe *any* fully definite view to him concerning the eligibility of comparison cases. He was, after all, still an undergraduate, and he was making most of this up *ex nihilo*.

5. Transitivity

The move that Lewis really needs to make in order to deal with pre-emption—or at any rate 'early' pre-emption—is the move he eventually makes in his 1973b. There he gives up on the claim that counterfactual dependence is necessary for causation, and replaces it with the claim that counterfactual dependence is necessary for causal *dependence* and causation is the ancestral of causal dependence. (As noted above, he still had not hit on his solution when he wrote the Gavin David Young Lectures, some time between 1968 and 1970.) This absolves him of the need to insist, in the two-drug case just discussed, that there is a proper comparison case to serve as an alternative to *A*—namely the case where Drug *C* is somehow neutralised, but not by Drug *A*—to make it true that *B* (the recovery) counterfactually depends on *A*. He can instead simply pick a suitable intermediary *D* between *A* and *B* such that *D* counterfactually depends on *A*, and *B* in turn counterfactually depends on *D*. (Take a time after Drug *C* is actually neutralised. At that time, Drug *A* is in the patient's system; call

that event *D*. Had *A* not happened, neither would *D*. And if *D* hadn't happened—given that Drug *C* was in fact already neutralised by *A* at that point—neither would *B*.)

That solution of course requires that causation is transitive—a position that Lewis never abandoned despite what many other authors have taken to be compelling counterexamples. As Lewis says in 'Causation as Influence':

> In rejecting the counterexamples . . . I think I am doing what historians do. They trace causal chains, and, without more ado, they conclude that what comes at the end of the chain was caused by what went before. If they did not, they could say little about historical causation; because, over intervals of any length, historical counterfactuals become so very speculative that nothing much can be known about the dependence of any event on its causal ancestors. (2000: 195)

Compare his remarks in Paper 3:

> There are many other things that can be wrong with a particular causal claim, making it uninformative, ridiculous, irrelevant, unfit to be mentioned, something we would never say, but not strictly *false*. I think these should not be considered part of the logical structure of 'cause'. It is a feature of our concept of cause that we can draw long and tortuous causal chains starting with a trivial event in the distant past and ending with something important in the present. It seems to me that a correct account of 'cause' should have a place for long causal chains, and should not even consider them as non-standard cases of 'cause.' (1961/2022: XX)

Lewis is defending transitivity in both of these passages. However, I suggest that the worries underlying the two passages are very different, concerning two different kinds of alleged counterexample.

In 'Causation as Influence' Lewis is explicitly concerned with (alleged) threat/aversion counterexamples, that is, cases with the following basic structure: *c* poses some threat to the occurrence of *e*; *d* averts the threat; *e* follows as a result. (I have the upper hand in our game of chess, but your clever move (*c*) threatens my queen. I manage to move her out of danger (*d*) and go on to win the game (*e*) as expected.) So we have a chain of counterfactual dependence—of *d* on *c*, and *e* on *d*—but no counterfactual dependence of *e* on *c* (if you hadn't threatened my queen I still would have won); and, it is alleged by many authors,

intuitively no causation of *e* by *c* either (2000: 194–5). Lewis claims that such cases are no threat to transitivity because it's perfectly acceptable to insist that *c* is, in fact, a cause of *e.* Counterfactual dependence is, after all, not necessary for causation: that's baked into the distinction between causation and causal dependence.

The implicit target in Paper 3, however, is most likely a case that Hart and Honoré bring up as a *prima facie* objection to the *sine qua non* view: if causation is a matter of counterfactual dependence, then a person's birth is amongst the causes of their injury.[14] Lewis is implicitly accepting the consequent of this conditional but denying that it is problematic: it may be inappropriate—and certainly in a legal context it would be wholly irrelevant—to *say* that the birth was a cause of the injury, but it is not *false*. The concern of the 1961 paper, then—unlike the challenge raised by threat/aversion cases—is a challenge to the claim that counterfactual dependence (between distinct 'situations') is *sufficient* for causation. So, despite his avowal of transitivity in that passage, the *problem* of transitivity, as per threat/aversion cases, isn't the problem he's addressing at all.

So why is Lewis so much as mentioning 'long and tortuous causal chains'? Why doesn't he simply confidently assert that the victim's birth is a cause of their injury, since the latter clearly counterfactually depends on the former, and be done? This is the line he takes later on, for example in 'Causation as Influence':

> 'Counterfactual analysis of causation? Yeah, yeah, my birth is a cause of my death!' said the scoffer. His birth is indeed a cause of his death; but it is understandable that we seldom say so. The counterfactual dependence of his death on his birth is just too obvious to be worth mentioning. (2000: 196)

Some light can be shed on this question by his February 1961 letter to Scriven, where he says:

> Where the cause is very remote, I balk, but can be persuaded to swallow it with a '. . . well, *strictly*, I suppose so, but . . .' or a '. . . yes, in a sense'.

[14] Hart and Honoré themselves reject this counterexample: while it is indeed a case of counterfactual dependence without causation, the connection between the birth and the injury is, they claim, 'analytic' rather than 'factual' (since it is—allegedly—analytic that anyone who is injured, and hence alive, has previously been born), and a 'factual' connection is required for causation (1959: 108–9).

> My grandfather's birth caused you to get this letter. I balk especially in a legal context; I find it gets quite palatable in the train of thought: a thousand years ago, a leaf fell; the consequences ramified through the ages . . .; my grandfather was born; I wrote this letter, etc.; in a thousand years think what great changes in the world caused by the fall of a single leaf! (Lewis in Beebee and Fisher 2020a: 3)

The birth/injury case that Hart and Honoré are concerned with is, of course, analogous to the birth/letter example Lewis uses here: they are both cases where, in legal terms, an event in the distant past is judged to be too 'remote' to count as a legal cause. Lewis is here expressing some reticence about simply accepting the verdict of his own theory; no wonder, then, that he doesn't (as I suggested above that he could) simply endorse that verdict and be done. But how exactly is the appeal to the even lengthier causal chain in which his grandfather's birth and the receipt of the letter are embedded (correspondingly, the appeal to 'long and tortuous causal chains' in Paper 3) supposed to help smooth the waters? I suggest the idea is simply that making salient the links in the causal chain between cause and effect—and perhaps, as suggested in the letter, making salient some even lengthier chain within which that chain is embedded—diminishes the salience of the fact that the cause is remote from the effect, since none of the individual causal connections we are now focussing on are ones such that the cause is remote from the effect. And of course transitivity is what assures us that an event linked to a later event in the chain via a series of more contiguous effects will be a cause of that later event.[15]

What Lewis has not yet seen, in 1961, is that transitivity conflicts with the identification of causation with counterfactual dependence; and of course that conflict will turn out to be a feature rather than a bug, because it is only by giving up on identifying causation with counterfactual dependence (and maintaining transitivity) that he finds a (*prima facie*) satisfactory solution to the pre-emption problem.

6. From 1961 to 1973

[15] In fact, in the next letter, also to Scriven—dated 21 July 1961, after he has written the final version of his 1961 paper—he says: 'whether "cause" is a transitive relation between particular situations depends on the method of selecting comparison cases in the context in which transitivity is asserted' (Letter 2 in Beebee and Fisher 2020a: 5). This suggests that he thinks he might be able to account for his ambivalence in the grandfather case in the same kind of way as he does in the case of pre-emption—which is not, as we saw in §5, a very promising suggestion.

As noted in §1, in November 1972 Lewis described the soon-to-be-written 'Causation' as his 'oldest piece of unfinished business'. The only evidence that Lewis thought about causation after 1961 but before he started work on *Counterfactuals* comes from a fourth paper, 'Singular causal explanation and indeterminism', written in November 1963 as a term paper for Philosophy 254, Morton White's theory of knowledge seminar at Harvard. (White awarded it an 'A'.) There, Lewis's aim is to come up with a broadly Hempelian account of explanation that can accommodate, *inter alia*, indeterministically-caused events whose probability is lower than the alternative. (My pressing the button makes it 75% likely that a light will flash and 25% likely that a bell will ring. I press the button and the bell rings. The ringing of the bell is explained, Lewis thinks, by the fact that I pressed the button.) Lewis notes that in such cases the putative explanans *C* is a *sine qua non* of the explanandum *A*, and suggests an analysis according to which the likelihood of *A* is higher than it would have been in the absence of *C*. Having read Goodman on counterfactuals by this point, he is aware of the cotenability problem. In this connection he notes that, in excising *C*, we 'cannot hold absolutely everything else constant' (1963: 20). But since we 'usually cannot tell exactly what somebody's supposition is, and so what is and is not cotenable with it', we 'have to fill it out from the context—if we can' (*ibid.*). Here again, then, we have the germ of the idea that, when it comes to counterfactual dependence, context is often required to resolve indeterminacy.

The 1963 paper doesn't add to or refine Lewis's earlier views on causation. Rather, it can be seen as a somewhat tentative attempt to promote the idea that even when it comes to the kinds of concern that philosophers were generally interested at the time—as discussed in §2 above—causation, conceived in terms of *sine qua non* or counterfactual dependence, might turn out to be rather more useful a notion than it was generally thought to be. We might surmise, then, that he had already decided that there was not much point in trying to develop a counterfactual theory of causation until he had a worked-out theory of counterfactuals to go with it.

As we saw in §4, however, by the time he wrote the Gavin David Young Lectures, some time between 1968 and 1970, he *had* been thinking about causation again. In Lecture 4 (as well as rehearsing the worry about pre-emption discussed in §4, which I'll get back to shortly), Lewis's main concern is with backwards causation in the context of time travel, where backwards time travel involves backwards causation and hence must, on his view of causation, involve backwards counterfactual dependence. He introduces his counterfactual analysis of causation in the context of a contrast between what he calls the 'existential

deductive-nomological analysis' (EDN) and the '*sine qua non* analysis', (SQN)—so roughly the contrast that framed the discussion of §2 above.

EDN, he says, 'goes badly wrong' (Lewis 1971/2022: XX): it can't deal with indeterministic causation, it suffers from what he later calls 'the problem of epiphenomena' (1973b: 170), and it implies that 'in all cases of overdetermination [i.e. redundant causation], even where one putative cause interferes with the action of the other [i.e. pre-emption], both are causes' (Lewis 1971/2022: XX).[16] He expresses the 'hunch' that 'EDN is wrong; a slight correction would make it right . . . and would make it equivalent to SQN' (*ibid.*). Unfortunately, he doesn't tell us what the 'slight correction' might be; and by the time he writes 'Causation' he has taken the view that its prospects are 'dark' (1973b: 557).

Lewis then turns his attention to SQN, noting that it 'has hitherto been unpopular because it involves counterfactual conditionals, and the analysis of these has seemed at least as obscure as that of causation' (Lewis 1971/2022: XX). At this point, he briefly sketches his theory of counterfactuals, which gets us to 'SQN+': 'event *C* is a cause of event *E* iff in all the worlds most closely resembling ours in which *C* is absent, *E* is also absent' (*ibid.*: XX). SQN+, he says, has no trouble with indeterministic causation and can solve the problem of epiphenomena. That leaves us with redundant causation, where SQN+ has the opposite problem to EDN: EDN's problem, in effect, is that it renders *both* potential causes as causes even in cases of pre-emption, since each is sufficient for the effect; SQN+'s problem is that it renders *neither* of them causes. Which leads to his discussion of pre-emption, already discussed in §4 above.[17]

As we saw in §4, at this point he is still sticking to the 'causation is counterfactual dependence' line and worrying about how to get the two to line up in pre-emption cases, where it's hard to see how to make it come out true that the effect counterfactually depends on the pre-empting cause. That position persists until at least May 1971 when he says:

[16] At this point, and in the letter to Woodruff discussed below, Lewis is clearly using 'overdetermination' to refer to redundant causation—that is, to both (symmetric) overdetermination and pre-emption. This makes sense because he still has some residual inclination to be ambivalent about whether (what he will shortly come to see as) the pre-empting cause *is* determinately a cause. He is clear, though, that the pre-empted alternative is *not* a cause, which is why redundant causation is especially problematic for EDN.

[17] Lewis goes on to apply SQN+ to backwards causation in time travel scenarios; I lack the space to discuss this here.

> My own tentative view, expounded in a section of *Counterfactuals,* is that if *c* and *e* are two events such that *e* occurred after *c* occurred, then *c* is a cause of *e* iff *c* did not occur > *e* did not occur. This makes for trouble, how serious I don't know, about overdetermination. (Letter 19 to Peter Woodruff, in Beebee and Fisher 2020a: 34)

So Lewis is still, as of May 1971, sticking to the old claim that counterfactual dependence is necessary for causation. And of course that *does* make trouble for overdetermination, or rather for redundant causation—and in particular for pre-emption.

It would appear that by the time he submitted *Counterfactuals* to the publisher in July 1972 he had decided that the trouble was in fact pretty serious, since the section he mentions in the passage above doesn't appear in the published version of *Counterfactuals*; indeed causation is not mentioned anywhere in the book.[18] It would appear, then, that Lewis did not hit on the idea of giving up the claim that counterfactual dependence is necessary for causation and instead making causation the ancestral of causal dependence—this being the solution to the 'early' pre-emption problem, which was the only problem of redundant causation that exercised him at the time—until he finally got back to his 'unfinished business' in late 1972 or early 1973.

The general picture all of this suggests, then, is that Lewis in effect parked the analysis of causation, having failed to get his 1961 paper published at the time, until at least the late 1960s, when he finally had a full-blown theory of counterfactuals to work with. Indeed we might speculate that Lewis's interest in developing such a theory was motivated by the desire to provide the foundations for a counterfactual analysis theory of causation. The theory of counterfactuals gave him most of the ingredients he needed—but not everything. All that remained was to figure out how to deal with pre-emption, which he duly did in the early 1970s.

Another important difference between Lewis's early and later views on causation concerns context dependence. As we saw in §4, there are two interpretations we might give of Lewis's claim that we sometimes have a 'choice' of comparison cases. According to the first, we have such a choice (which can be resolved either by context or by explicit contrast) just when it is indeterminate what would have happened had the cause not occurred. According to the second, early Lewis is a contrastivist about counterfactual dependence:

---

[18] We might hypothesise that what Lewis says in that excised section is broadly in line with what he says in the Gavin David Young Lectures.

whether or not *E* counterfactually depends on *C* can depend on our choice of contrast case even when there is no such indeterminacy—that is, even when one contrast is (to put it anachronistically) determinately more of a departure from actuality than the other.

The context-dependence of counterfactual dependence is also a feature of Lewis's 1973a view of counterfactuals, in that comparative similarity is vague and the vagueness is to be resolved by context. And, as we saw earlier, in his Gavin David Young Lectures he is still toying with the idea that the context-dependence of counterfactuals delivers context-dependence for causation itself. Indeed he explicitly claims as much in 'Causation', saying: 'The vagueness of similarity does infect causation, and no correct analysis can deny it' (1973b: 163).

Strikingly, however, he shows no interest in that paper in explaining *how* the vagueness of similarity infect causation, and hence why 'no correct analysis can deny it'. Neither there nor in later papers does he present any putative cases where different and equally legitimate resolutions of vagueness might deliver different but equally legitimate causal judgements in different contexts. In 'Counterfactual dependence and time's arrow', he gives us a list of priorities that constitute what he calls the 'standard resolution of vagueness' (1979: 47). And he says just afterwards: 'Plenty of unresolved vagueness remains, of course' (1979: 48). But again there is no suggestion that such unresolved vagueness will or might deliver any conflicting judgements when it comes to the counterfactuals involved in causation. His mature view, then, seems to be this: first, while vagueness is indeed an ineliminable feature of counterfactuals, the 'standard' resolution is not merely standard but obligatory in contexts where causal judgements are in play; and, second, the standard resolution delivers *enough* determinacy to prevent any residual vagueness from infecting our causal judgements.

Later Lewis also clearly has no time for the idea that causation is context-dependent in the way that contemporary contrastivists (or indeed causal contextualists more generally) think it is. In his 1973b he explains that he wants to capture 'a broad and nondiscriminatory concept of causation' (1973b: 559)—one that is insensitive to (*inter alia*) contextual salience. Causal contextualists typically motivate their views by appealing to cases where context makes a difference to whether or not a given causal claim is acceptable. For example, the claim that the presence of oxygen caused a forest fire would be unacceptable to the forest rangers but acceptable to visiting Venusians (Schaffer 2005: 37–8); and the claim that the gardener's failure to water my plants caused them to die is acceptable but the claim that the Queen's failure to water them caused them to die is unacceptable—despite the fact that their death counterfactually depends equally on both. Contextualists explain such differences in

acceptability in terms of truth and falsity, and hence build contextual salience into their accounts of the truth conditions of causal claims. But Lewis's view on such cases is clearly that the role of contextual salience is merely pragmatic: context determines which causes are worth mentioning, and not which causes there *are.*[19]

In other words, according to later Lewis, neither kind of context-dependence—neither the kind that resolves the vagueness of the similarity relation, nor the kind appealed to by contextualists and contrastivists about causation—has any role to play in the correct analysis of causation.[20] We can only speculate about what prompted this shift away from the idea that our causal judgements are context-dependent. Perhaps the idea was only there to start with because Lewis thought it was the key to solving the pre-emption problem, which he no longer needed once he'd found a better approach. Perhaps his intuitions about pre-emption cases firmed up in the intervening years, so that he simply no longer felt any ambivalence about whether or not pre-empting causes really are causes. Or perhaps he was independently convinced of the need to capture 'a broad and nondiscriminatory concept of causation' (1973b: 559).

It seems appropriate to end by indulging in a little counterfactual history: how might things have gone differently for the metaphysics of causation over the last 60 years if *Analysis* or *The Journal of Philosophy* had seen fit to publish Lewis's 1961 paper? I suspect that it wouldn't have made much difference. No remotely viable theory of counterfactuals was available in the early 1960s, and counterfactuals were generally not regarded as being especially central to philosophical analysis, doubtless due to the widespread suspicion of modality that was inherent in empiricist thinking generally and consolidated by Quine. Moreover, the dominant interest in induction and confirmation—and, again, empiricist scruples—meant that particular causation was simply not a topic that many philosophers were interested in. So the 1961 paper, had it been published, might well have languished in obscurity. By 1973, however, the sands had shifted. Both Lewis and Stalnaker (1968) had

[19] See for example Lewis 1986b, §1; 2004, 196.

[20] Actually context-dependence does re-enter the picture in Lewis's 'Causation as influence' (2000). On that view, even a pre-empted 'spurious cause' such as Billy's rock-throwing (pre-empted by Suzy's slightly earlier or faster throw) might make a tiny difference to the shattering of the window via its gravitational effects. We are entitled to ignore this tiny difference, Lewis thinks; but 'if for some strange reason we did attend to these negligible differences', we would 'then put ourselves in an unusual context where it was right, not wrong, to count all the things that make negligible differences as joint causes of the effect' (2000: 189). This kind of context-dependence, however, has nothing to do with either the vagueness of the similarity relation or the salience of contrasts.

developed new theories of counterfactuals.[21] Metaphysics was emerging once again as a popular field of enquiry thanks to the work of Smart, Armstrong, Kripke and others. Correspondingly, causation—and specifically 'particular' causation—was increasingly being seen as a topic of legitimate metaphysical interest in its own right, and was finally getting some attention (e.g. from Mackie 1965, Ducasse 1965, and Davidson 1967). Lewis's original counterfactual analysis was too far ahead of its time; it needed to wait for the philosophical world—including Lewis himself, via his theory of counterfactuals—to catch up.[22]

Bibliography

American Law Institute. 1965. *Restatement (Second) of Torts.*

Beebee, Helen. 2006. *Hume on Causation.* Abingdon: Routledge.

Beebee, Helen and A.R.J. Fisher, eds. 2020a. *Philosophical Letters of David K. Lewis, Vol. I: Causation, Modality, Ontology.* Oxford: Oxford University Press.

Beebee, Helen and A.R.J. Fisher, eds. 2020b. *Philosophical Letters of David K. Lewis, Vol. II: Mind, Language, Epistemology.* Oxford: Oxford University Press.

Bernstein, Sara. 2019. Lewis's Theories of Causation and their Influence. In *The Cambridge History of Philosophy, 1945–2015,* eds. K. Becker and I. Thomson. Cambridge: Cambridge University Press, 160–70.

Bingham, Joseph W. 1909. Some Suggestions concerning "Legal Cause" at Common Law. *Columbia Law Review* 9: 16–37.

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

Chisholm, Roderick. 1946. The Contrary-to-Fact Conditional. *Mind* 55: 289–307.

Chisholm, Roderick. 1955. Law Statements and Counterfactual Inference. *Analysis* 15: 97–105.

[21] Lewis and Stalnaker arrived at their theories independently; both are also similar to an account proposed by Timothy Sprigge (1970). See Letter 7. To Robert C. Stalnaker, 31 May 1968 and Letter 82. To T.L.S. Sprigge, 15 November 1993, both in Beebee and Fisher 2020a: 16 and 162.

[22] Thanks to the audience at the Arché metaphysics and logic seminar at St. Andrews for discussion. And undying gratitude to Anthony Fisher—not just for his help with this chapter but, more generally, for the enormous amount of work and enthusiasm he put into the project from which this chapter arose (*The Age of Metaphysical Revolution: David Lewis and His Place in the History of Analytic Philosophy*), his encyclopaedic knowledge of all things Lewis, and his generosity in sharing it. I also acknowledge the generous support of the AHRC for its funding of the project (grant no. AH/N004000/1).

Davidson, Donald. 1967. Causal Relations. *The Journal of Philosophy* 64: 691–703.

Ducasse, C.J. 1965. Causation: Perceivable? Or Only Inferred? *Philosophy and Phenomenological Research* 26: 173–9.

Ewing, A.C. 1932–1933. A Defence of Causality. *Proceedings of the Aristotelian Society* 33: 95–128.

Feigl, Herbert. 1953. Notes on Causality. In *Readings in the Philosophy of Science,* eds. H. Feigl and M. Brodbeck. New York: Appleton-Century-Crofts, 408–18.

Garrett, Don. 1997. *Cognition and Commitment in Hume's Philosophy.* New York: Oxford University Press.

Goodman, Nelson. 1947. *Fact, Fiction, and Forecast.* Cambridge, MA: Harvard University Press.

Hanson, R.N. 1955. Causal Chains. *Mind* 64: 289–311.

Hart, H.L.A. and Tony Honoré. 1959. *Causation in the Law.* Oxford: Clarendon Press.

Hempel, Carl. 1945. Studies in the Logic of Confirmation. *Mind* 54: 1–26 (Part I); 97–121 (Part II).

Janssen-Lauret, Frederique and Fraser MacBride, eds. 2022. *David Lewis: Philosophical Manuscripts*. Oxford: Oxford University Press.

Lewis, David K. 1958. On Causality and Natural Laws. Unpublished m/s. David Lewis Papers, C1520, Manuscripts Division, Department of Special Collections, Princeton University Library.

Lewis, David K. 1959–60. Cause. Unpublished m/s. David Lewis Papers, C1520, Manuscripts Division, Department of Special Collections, Princeton University Library.

Lewis, David K. 1961/2022. Particular and General Causal Claims. In Janssen-Lauret and MacBride 2022, XX–XX.

Lewis, David K. 1963. Singular Causal Explanation and Indeterminism. David Lewis Papers, C1520, Manuscripts Division, Department of Special Collections, Princeton University Library.

Lewis, David K. 1971/2022. The Paradoxes of Time Travel: The Gavin David Young Lectures at the University of Adelaide, 1971. In Janssen-Lauret and MacBride 2022, XX–XX.

Lewis, David K. 1973a. *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, David K. 1973b. Causation. *The Journal of Philosophy* 70: 556–67.

Lewis, David K. 1979. Counterfactual Dependence and Time's Arrow. *Noûs* 13: 455–76.

Lewis, David K. 1986a. Postscripts to ‘Causation’. In his *Philosophical Papers II*. Oxford: Oxford University Press, 172–213.

Lewis, David K. 1986b. Causal Explanation. In his *Philosophical Papers II*. Oxford: Oxford University Press, 214–240.

Lewis, David K. 2000. Causation as Influence. *The Journal of Philosophy* 97: 182–97.

Lewis, David K. 2004. Void and Object. In Collins, Hall and Paul 2004, 277–90.

Mackie, J.L. 1965. Causes and Conditions. *American Philosophical Quarterly* 2: 245–64.

Maslen, Cei. 2004. Causes, Contrasts, and the Nontransitivity of Causation. In Collins, Hall and Paul 2004, 341–58.

Northcott, Robert. 2008. Causation and Contrast Classes. *Philosophical Studies* 139: 111–23.

Popper, Karl. 1949. A Note on Natural Laws and So-Called ‘Contrary-to-Fact Conditionals’. *Mind* 58: 62–6.

Popper, Karl. 1959. *The Logic of Scientific Discovery.* London: Hutchinson.

Rue, John D. 2003. Returning to the Roots of the Bramble Bush: The ‘But For’ Test Regains Primacy in Causal Analysis in the American Law Institute’s Proposed Restatement (Third) of Torts. *Fordham Law Review* 71: 2679–734.

Russell, Bertrand. 1912–1913. On the Notion of Cause. *Proceedings of the Aristotelian Society* 13: 1–26.

Schaffer, Jonathan. 2005. Contrastive Causation. *The Philosophical Review* 114: 297–328.

Schneider, E.F. 1953. Recent Discussion of Subjunctive Conditionals. *The Review of Metaphysics* 6: 623–49.

Smith, Norman Kemp. 1905. The Naturalism of Hume (I). *Mind* 54: 149–73.

Smith, Norman Kemp. 1941. *The Philosophy of David Hume.* London: Macmillan.

Sprigge, T.L.S. 1970. *Facts, Worlds and Beliefs.* London: Routledge & Kegan Paul.

Stalnaker, Robert. 1968. A Theory of Conditionals. In *Studies in Logical Theory,* ed. Nicholas Rescher. Oxford: Basil Blackwell, 98–112.