Twenty-One Causation [*]

David Lewis

Hume defined causation twice over. He wrote "we may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second never had existed*." [1]

[1] *An Enquiry concerning Human Understanding*, Section VII.

Descendants of Hume's first definition still dominate the philosophy of causation: a casual succession is supposed to be a succession that instantiates a regularity. To be sure, there have been improvements. Nowadays we try to distinguish the regularities that count—the "causal laws"—from mere accidental regularities of succession. We subsume causes and effects under regularities by means of descriptions they satisfy, not by over-all similarity. And we allow a cause to be only one indispensable part, not the whole, of the total situation that is followed by the effect in accordance with a law. In present-day regularity analyses, a cause is defined (roughly) as any member of any minimal set of actual conditions that are jointly sufficient, given the laws, for the existence of the effect.

More precisely, let C be the proposition that c exists (or occurs) and

let E be the proposition that e exists. Then c causes e, according to a typical regularity analysis, [2]

[2] Not one that has been proposed by any actual author in just this form, so far as I know.

iff (1) C and E are true; and (2) for some nonempty set $\mathcal{L}$ of true law-propositions and some set $\mathcal{F}$ of true propositions of particular fact, $\mathcal{L}$ and $\mathcal{F}$ jointly imply $C \supset E$, although $\mathcal{L}$ and $\mathcal{F}$ jointly do not imply E and $\mathcal{F}$ alone does not imply $C \supset E$. [3]

[3] I identify a *proposition*, as is becoming usual, with the set of possible worlds where it is true. It is not a linguistic entity. Truth-functional operations on propositions are the appropriate Boolean operations on sets of worlds; logical relations among propositions are relations of inclusion, overlap, etc. among sets. A sentence of a language *expresses* a proposition iff the sentence and the proposition are true at exactly the same worlds. No ordinary language will provide sentences to express all propositions; there will not be enough sentences to go around.

Much needs doing, and much has been done, to turn definitions like this one into defensible analyses. Many problems have been overcome. Others remain: in particular, regularity analyses tend to confuse causation itself with various other causal relations. If c belongs to a minimal set of conditions jointly sufficient for e, given the laws, then c may well be a genuine cause of e. But c might rather be an effect of e: one which could not, given the laws and some of the actual circumstances, have occurred otherwise than by being caused by e. Or c might be an epiphenomenon of the causal history of e: a more or less inefficacious effect of some genuine cause of e. Or c might be a preempted potential cause of e: something that did not cause e, but that would have done so in the absence of whatever really did cause e.

It remains to be seen whether any regularity analysis can succeed in distinguishing genuine causes from effects, epiphenomena, and preempted potential causes—and whether it can succeed without falling victim to worse problems, without piling on the epicycles, and without departing from the fundamental idea that causation is instantiation of regularities. I have no proof that regularity analyses are beyond repair, nor any space to review the repairs that have been tried. Suffice it to say that the prospects look dark. I think it is time to give up and try something else.

A promising alternative is not far to seek. Hume's "other words"—that if the cause had not been, the effect never had existed—are no mere restatement of his first definition. They propose something altogether different: a counterfactual analysis of causation.

The proposal has not been well received. True, we do know that causation has something or other to do with counterfactuals. We think

of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects—some of them, at least, and usually all—would have been absent as well. Yet it is one thing to mention these platitudes now and again, and another thing to rest an analysis on them. That has not seemed worth while. [4]

[4] One exception: Ardon Lyon, "Causality," *British Journal for the Philosophy of Science* , XVIII, 1 (May 1967): 1–20.

We have learned all too well that counterfactuals are ill understood, wherefore it did not seem that much understanding could be gained by using them to analyze causation or anything else. Pending a better understanding of counterfactuals, moreover, we had no way to fight seeming counterexamples to a counterfactual analysis.

But counterfactuals need not remain ill understood, I claim, unless we cling to false preconceptions about what it would be like to understand them. Must an adequate understanding make no reference to unactualized possibilities? Must it assign sharply determinate truth conditions? Must it connect counterfactuals rigidly to covering laws? Then none will be forthcoming. So much the worse for those standards of adequacy. Why not take counterfactuals at face value: as statements about possible alternatives to the actual situation, somewhat vaguely specified, in which the actual laws may or may not remain intact? There are now several such treatments of counterfactuals, differing only in details. [5]

[5] See, for instance, Robert Stalnaker, "A Theory of Conditionals," in Nicholas Rescher, ed., *Studies in Logical Theory* (Oxford: Blackwell, 1968); and my *Counterfactuals* (Oxford: Blackwell, 1973).

If they are right, then sound foundations have been laid for analyses that use counterfactuals.

In this paper, I shall state a counterfactual analysis, not very different from Hume's second definition, of some sorts of causation. Then I shall try to show how this analysis works to distinguish genuine causes from effects, epiphenomena, and preempted potential causes.

My discussion will be incomplete in at least four ways. Explicit preliminary settings-aside may prevent confusion.

1. I shall confine myself to causation among *events*, in the everyday sense of the word: flashes, battles, conversations, impacts, strolls, deaths, touchdowns, falls, kisses, and the like. Not that events are the only things that can cause or be caused; but I have no full list of the others, and no good umbrella-term to cover them all.

2. My analysis is meant to apply to causation in particular cases. It

is not an analysis of causal generalizations. Presumably those are quantified statements involving causation among particular events (or non-events), but it turns out not to be easy to match up the causal generalizations of natural language with the available quantified forms. A sentence of the form "c-events cause E-events," for instance, can mean any of

   (a) For some c in C and some e in E, c causes e.
   (b) For every e in E, there is some c in C such that c causes e.
   (c) For every c in C, there is some e in E such that c causes e.
       not to mention further ambiguities. Worse still, "Only c-events cause E-events" ought to mean
   (d) For every c, if there is some e in E such that c causes e, then c is in C.
       if "only" has its usual meaning. But no; it unambiguously means (b) instead! These problems are not about causation, but about our idioms of quantification.

3. We sometimes single out one among all the causes of some event and call it "the" cause, as if there were no others. Or we single out a few as the "causes," calling the rest mere "causal factors" or "causal conditions." Or we speak of the "decisive" or "real" or "principal" cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. [6]

   [6] Except that Morton G. White's discussion of causal selection, in *Foundations of Historical Knowledge* (New York: Harper & Row, 1965), pp. 105–181, would meet my needs, despite the fact that it is based on a regularity analysis.

   I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation.

4. I shall be content, for now, if I can give an analysis of causation that works properly under determinism. By determinism I do not mean any thesis of universal causation, or universal predictability-in-principle, but rather this: the prevailing laws of nature are such that there do not exist any two possible worlds which are exactly alike up to some time, which differ thereafter, and in which those laws are never violated. Perhaps by ignoring indeterminism I squander the most striking advantage of a counterfactual analysis over a regularity analysis:

that it allows undetermined events to be caused. [7]

[7] That this ought to be allowed is argued in G. E. M. Anscombe, *Causality and Determination: An Inaugural Lecture* (Cambridge: University Press, 1971); and in Fred Dretske and Aaron Snyder, "Causal Irregularity," *Philosophy of Science* , XXXIX, 1 (March 1972): 69–71.

I fear, however, that my present analysis cannot yet cope with all varieties of causation under indeterminism. The needed repair would take us too far into disputed questions about the foundations of probability.

Comparative Similarity

To begin, I take as primitive a relation of *comparative over-all similarity* among possible worlds. We may say that one world is *closer to actuality* than another if the first resembles our actual world more than the second does, taking account of all the respects of similarity and difference and balancing them off one against another.

(More generally, an arbitrary world w can play the role of our actual world. In speaking of our actual world without knowing just which world is ours, I am in effect generalizing over all worlds. We really need a three-place relation: world w_1 is closer to world w than world w_2 is. I shall henceforth leave this generality tacit.)

I have not said just how to balance the respects of comparison against each other, so I have not said just what our relation of comparative similarity is to be. Not for nothing did I call it primitive. But I have said what *sort* of relation it is, and we are familiar with relations of that sort. We do make judgments of comparative overall similarity—of people, for instance—by balancing off many respects of similarity and difference. Often our mutual expectations about the weighting factors are definite and accurate enough to permit communication. I shall have more to say later about the way the balance must go in particular cases to make my analysis work. But the vagueness of over-all similarity will not be entirely resolved. Nor should it be. The vagueness of similarity does infect causation, and no correct analysis can deny it.

The respects of similarity and difference that enter into the over-all similarity of worlds are many and varied. In particular, similarities in matters of particular fact trade off against similarities of law. The prevailing laws of nature are important to the character of a world; so similarities of law are weighty. Weighty, but not sacred. We should not take it for granted that a world that conforms perfectly to our actual

laws is *ipso facto* closer to actuality than any world where those laws are violated in any way at all. It depends on the nature and extent of the violation, on the place of the violated laws in the total system of laws of nature, and on the countervailing similarities and differences in other respects. Likewise, similarities or differences of particular fact may be more or less weighty, depending on their nature and extent. Comprehensive and exact similarities of particular fact throughout large spatio-temporal regions seem to have special weight. It may be worth a small miracle to prolong or expand a region of perfect match.

Our relation of comparative similarity should meet two formal constraints. (1) It should be a weak ordering of the worlds: an ordering in which ties are permitted, but any two worlds are comparable. (2) Our actual world should be closest to actuality, resembling itself more than any other world resembles it. We do *not* impose the further constraint that for any set A of worlds there is a unique closest A-world, or even a set of A-worlds tied for closest. Why not an infinite sequence of closer and closer A-worlds, but no closest?

Counterfactuals and Counterfactual Dependence

Given any two propositions A and C, we have their *counterfactual* $A \,\square\!\!\rightarrow C$: the proposition that if A were true, then C would also be true. The operation $\square\!\!\rightarrow$ is defined by a rule of truth, as follows. $A \,\square\!\!\rightarrow C$ is true (at a world w) iff either (1) there are no possible A-worlds (in which case $A \,\square\!\!\rightarrow C$ is *vacuous*), or (2) some A-world where C holds is closer (to w) than is any A-world where C does not hold. In other words, a counterfactual is nonvacuously true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent.

We did not assume that there must always be one or more closest A-worlds. But if there are, we can simplify: $A \,\square\!\!\rightarrow C$ is nonvacuously true iff C holds at all the closest A-worlds.

We have not presupposed that A is false. If A is true, then our actual world is the closest A-world, so $A \,\square\!\!\rightarrow C$ is true iff C is. Hence $A \,\square\!\!\rightarrow C$ implies the material conditional $A \supset C$; and A and C jointly imply $A \,\square\!\!\rightarrow C$.

Let $A_1, A_2, \ldots$ be a family of possible propositions, no two of which are compossible; let $C_1, C_2, \ldots$ be another such family (of

equal size). Then if all the counterfactuals $A_1 \,\square\!\!\rightarrow C_1$, $A_2 \,\square\!\!\rightarrow C_2$, . . . between corresponding propositions in the two families are true, we shall say that the C's *depend counterfactually* on the A's. We can say it like this in ordinary language: whether C_1 or C_2 or . . . depends (counterfactually) on whether A_1 or A_2 or . . . .

Counterfactual dependence between large families of alternatives is characteristic of processes of measurement, perception, or control. Let R_1, R_2, . . . be propositions specifying the alternative readings of a certain barometer at a certain time. Let P_1, P_2, . . . specify the corresponding pressures of the surrounding air. Then, if the barometer is working properly to measure the pressure, the R's must depend counterfactually on the P's. As we say it: the reading depends on the pressure. Likewise, if I am seeing at a certain time, then my visual impressions must depend counterfactually, over a wide range of alternative possibilities, on the scene before my eyes. And if I am in control over what happens in some respect, then there must be a double counterfactual dependence, again over some fairly wide range of alternatives. The outcome depends on what I do, and that in turn depends on which outcome I want. [8]

[8] Analyses in terms of counterfactual dependence are found in two papers of Alvin I. Goldman: "Toward a Theory of Social Power," *Philosophical Studies* , XXIII (1972): 221–268; and "Discrimination and Perceptual Knowledge," presented at the 1972 Chapel Hill Colloquium.

Causal Dependence Among Events

If a family C_1, C_2, . . . depends counterfactually on a family A_1, A_2, . . . in the sense just explained, we will ordinarily be willing to speak also of causal dependence. We say, for instance, that the barometer reading depends causally on the pressure, that my visual impressions depend causally on the scene before my eyes, or that the outcome of something under my control depends causally on what I do. But there are exceptions. Let G_1, G_2, . . . be alternative possible laws of gravitation, differing in the value of some numerical constant. Let M_1, M_2, . . . be suitable alternative laws of planetary motion. Then the M's may depend counterfactually on the G's, but we would not call this dependence causal. Such exceptions as this, however, do not involve any sort of dependence among distinct particular events. The hope remains that causal dependence among events, at least, may be analyzed simply as counterfactual dependence.

end p.165

We have spoken thus far of counterfactual dependence among propositions, not among events. Whatever particular events may be, presumably they are not propositions. But that is no problem, since they can at least be paired with propositions. To any possible event e, there corresponds the proposition $O(e)$ that holds at all and only those worlds where e occurs. This $O(e)$ is the proposition that e occurs. [9]

[9] Beware: if we refer to a particular event e by means of some description that e satisfies, then we must take care not to confuse $O(e)$, the proposition that e itself occurs, with the different proposition that some event or other occurs which satisfies the description. It is a contingent matter, in general, what events satisfy what descriptions. Let e be the death of Socrates—the death he actually died, to be distinguished from all the different deaths he might have died instead. Suppose that Socrates had fled, only to be eaten by a lion. Then e would not have occurred, and $O(e)$ would have been false; but a different event would have satisfied the description "the death of Socrates" that I used to refer to e. Or suppose that Socrates had lived and died just as he actually did, and afterwards was resurrected and killed again and resurrected again, and finally became immortal. Then no event would have satisfied the description. (Even if the temporary deaths are real deaths, neither of the two can be *the* death.) But e would have occurred, and $O(e)$ would have been true. Call a description of an event e *rigid* iff (1) nothing but e could possibly satisfy it, and (2) e could not possibly occur without satisfying it. I have claimed that even such commonplace descriptions as "the death of Socrates" are nonrigid, and in fact I think that rigid descriptions of events are hard to find. That would be a problem for anyone who needed to associate with every possible event e a sentence $\Phi (e)$ true at all and only those worlds where e occurs. But we need no such sentences —only propositions, which may or may not have expressions in our language.

(If no two events occur at exactly the same worlds—if, that is, there are no absolutely necessary connections between distinct events —we may add that this correspondence of events and propositions is one to one.) Counterfactual dependence among events is simply counterfactual dependence among the corresponding propositions.

Let c_1, c_2, . . . and e_1, e_2, . . . be distinct possible events such that no two of the c's and no two of the e's are compossible. Then I say that the family e_1, e_2, . . . of events *depends causally* on the family c_1, c_2, . . . iff the family $O(e_1)$, $O(e_2)$, . . . of propositions depends counterfactually on the family $O(c_1)$, $O(c_2)$, ... As we say it: whether e_1 or e_2 or . . . occurs depends on whether c_1 or c_2 or . . . occurs.

We can also define a relation of dependence among single events rather than families. Let c and e be two distinct possible particular events. Then e *depends causally* on c iff the family $O(e)$, $\sim O(e)$ depends counterfactually on the family $O(c)$, $\sim O(c)$. As we say it: whether e occurs or not depends on whether c occurs or not. The dependence consists in the truth of two counterfactuals:

end p.166

$O(c) \,\square\!\!\rightarrow O(e)$ and $\sim O(c) \,\square\!\!\rightarrow \sim O(e)$. There are two cases. If c and e do not actually occur, then the second counterfactual is automatically true because its antecedent and consequent are true: so e depends causally on c iff the first counterfactual holds. That is, iff e would have occurred if c had occurred. But if c and e are actual events, then it is the first counterfactual that is automatically true.

Then e depends causally on c iff, if c had not been, e never had existed. I take Hume's second definition as my definition not of causation itself, but of causal dependence among actual events.

Causation

Causal dependence among actual events implies causation. If c and e are two actual events such that e would not have occurred without c, then c is a cause of e. But I reject the converse. Causation must always be transitive; causal dependence may not be; so there can be causation without causal dependence. Let c, d, and e be three actual events such that d would not have occurred without c and e would not have occurred without d. Then c is a cause of e even if e would still have occurred (otherwise caused) without c.

We extend causal dependence to a transitive relation in the usual way. Let c, d, e, . . . be a finite sequence of actual particular events such that d depends causally on c, e on d, and so on throughout. Then this sequence is a *causal chain*. Finally, one event is a *cause* of another iff there exists a causal chain leading from the first to the second. This completes my counterfactual analysis of causation.

Counterfactual Versus Nomic Dependence

It is essential to distinguish counterfactual and causal dependence from what I shall call *nomic dependence*. The family C_1, C_2, . . . of propositions depends nomically on the family A_1, A_2, . . . iff there are a nonempty set $\mathcal{L}$ of true law-propositions and a set $\mathcal{F}$ of true propositions of particular fact such that $\mathcal{L}$ and $\mathcal{F}$ jointly imply (but $\mathcal{F}$ alone does not imply) all the material conditionals $A_1 \supset C_1$, $A_2 \supset C_2$, . . . between the corresponding propositions in the two families. (Recall that these same material conditionals are implied by the counterfactuals that would comprise a counterfactual dependence.) We shall say

end p.167

also that the nomic dependence holds *in virtue of* the premise sets $\mathcal{L}$ and $\mathcal{F}$.

Nomic and counterfactual dependence are related as follows. Say that a proposition B is *counterfactually independent* of the family A_1, A_2, . . . of alternatives iff B would hold no matter which of the A's were true—that is, iff the counterfactuals $A_1 \ \square\!\!\rightarrow B$, $A_2 \ \square\!\!\rightarrow B$, . . . all hold. If the C's depend nomically on the A's in virtue of the premise sets $\mathcal{L}$ and $\mathcal{F}$, and if in addition (all members of) $\mathcal{L}$ and $\mathcal{F}$ are counterfactually independent of the A's, then it follows that the C's depend counterfactually on the A's. In that case, we may regard the nomic dependence in virtue of $\mathcal{L}$ and $\mathcal{F}$ as explaining the counterfactual dependence. Often, perhaps always, counterfactual dependences may be thus explained. But the requirement of counterfactual independence is indispensable. Unless $\mathcal{L}$ and $\mathcal{F}$ meet that requirement, nomic dependence in virtue of $\mathcal{L}$ and $\mathcal{F}$ does not imply counterfactual dependence, and, if there is counterfactual dependence anyway, does not explain it.

Nomic dependence is reversible, in the following sense. If the family C_1, C_2, . . . depends nomically on the family A_1, A_2, . . . in virtue of $\mathcal{L}$ and $\mathcal{F}$, then also A_1, A_2, . . . depends nomically on the family AC_1, AC_2, . . . , in virtue of $\mathcal{L}$ and $\mathcal{F}$, where A is the disjunction $A_1 \vee A_2 \vee$ . . . . Is counterfactual dependence likewise reversible? That does not follow. For, even if $\mathcal{L}$ and $\mathcal{F}$ are independent of A_1, A_2, . . . and hence establish the counterfactual dependence of the C's on the A's, still they may fail to be independent of AC_1, AC_2, . . . , and hence may fail to establish the reverse counterfactual dependence of the A's on the AC's. Irreversible counterfactual dependence is shown below: @ is our actual world, the dots are the other worlds, and distance on the page represents similarity "distance".
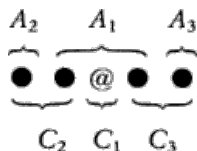
The counterfactuals $A_1 \ \square\!\!\rightarrow C_1$, $A_2 \ \square\!\!\rightarrow C_2$, and $A_3 \ \square\!\!\rightarrow C_3$ hold at the actual world; wherefore the C's depend on the A's. But we do not have the reverse dependence of the A's on the AC's, since instead of the needed $AC_2 \ \square\!\!\rightarrow A_2$ and $AC_3 \ \square\!\!\rightarrow A_3$ we have $AC_2 \ \square\!\!\rightarrow A_1$ and $AC_3 \ \square\!\!\rightarrow A_1$.

Just such irreversibility is commonplace. The barometer reading

end p.168

depends counterfactually on the pressure—that is as clear-cut as counterfactuals ever get—but does the pressure depend

counterfactually on the reading? If the reading had been higher, would the pressure have been higher? Or would the barometer have been malfunctioning? The second sounds better: a higher reading would have been an incorrect reading. To be sure, there are actual laws and circumstances that imply and explain the actual accuracy of the barometer, but these are no more sacred than the actual laws and circumstances that imply and explain the actual pressure. Less sacred, in fact. When something must give way to permit a higher reading, we find it less of a departure from actuality to hold the pressure fixed and sacrifice the accuracy, rather than vice versa. It is not hard to see why. The barometer, being more localized and more delicate than the weather, is more vulnerable to slight departures from actuality. [10]

[10] Granted, there are contexts or changes of wording that would incline us the other way. For some reason, "If the reading had been higher, that would have been because the pressure was higher" invites my assent more than "If the reading had been higher, the pressure would have been higher." The counterfactuals from readings to pressures are much less clear-cut than those from pressures to readings. But it is enough that some legitimate resolutions of vagueness give an irreversible dependence of readings on pressures. Those are the resolutions we want at present, even if they are not favored in all contexts.

We can now explain why regularity analyses of causation (among events, under determinism) work as well as they do. Suppose that event c causes event e according to the sample regularity analysis that I gave at the beginning of this paper, in virtue of premise sets $\mathcal{L}$ and $\mathcal{F}$. It follows that $\mathcal{L}$, $\mathcal{F}$ and $\sim O(c)$ jointly do not imply $O(e)$. Strengthen this: suppose further that they do imply $\sim O(e)$. If so, the family $O(e)$, $\sim O(e)$ depends nomically on the family $O(c)$, $\sim O(c)$ in virtue of $\mathcal{L}$ and $\mathcal{F}$. Add one more supposition: that $\mathcal{L}$ and $\mathcal{F}$ are counterfactually independent of $O(c)$, $\sim O(c)$. Then it follows according to my counterfactual analysis that e depends counterfactually and causally on c, and hence that c causes e. If I am right, the regularity analysis gives conditions that are almost but not quite sufficient for explicable causal dependence. That is not quite the same thing as causation; but causation without causal dependence is scarce, and if there is inexplicable causal dependence we are (understandably!) unaware of it. [11]

[11] I am not here proposing a repaired regularity analysis. The repaired analysis would gratuitously rule out inexplicable causal dependence, which seems bad. Nor would it be squarely in the tradition of regularity analyses any more. Too much else would have been added.

Effects and Epiphenomena

I return now to the problems I raised against regularity analyses, hoping to show that my counterfactual analysis can overcome them.

The *problem of effects*, as it confronts a counterfactual analysis, is as follows. Suppose that c causes a subsequent event e, and that e does not also cause c. (I do not rule out closed causal loops a priori, but this case is not to be one.) Suppose further that, given the laws and some of the actual circumstances, c could not have failed to cause e. It seems to follow that if the effect e had not occurred, then its cause c would not have occurred. We have a spurious reverse causal dependence of c on e, contradicting our supposition that e did not cause c.

The *problem of epiphenomena*, for a counterfactual analysis, is similar. Suppose that e is an epiphenomenal effect of a genuine cause c of an effect f. That is, c causes first e and then f, but e does not cause f. Suppose further that, given the laws and some of the actual circumstances, c could not have failed to cause e; and that, given the laws and others of the circumstances, f could not have been caused otherwise than by c. It seems to follow that if the epiphenomenon e had not occurred, then its cause c would not have occurred and the further effect f of that same cause would not have occurred either. We have a spurious causal dependence of f on e, contradicting our supposition that e did not cause f.

One might be tempted to solve the problem of effects by brute force: insert into the analysis a stipulation that a cause must always precede its effect (and perhaps a parallel stipulation for causal dependence). I reject this solution. (1) It is worthless against the closely related problem of epiphenomena, since the epiphenomenon e does precede its spurious effect f. (2) It rejects a priori certain legitimate physical hypotheses that posit backward or simultaneous causation. (3) It trivializes any theory that seeks to define the forward direction of time as the predominant direction of causation.

The proper solution to both problems, I think, is flatly to deny the counterfactuals that cause the trouble. If e had been absent, it is not that c would have been absent (and with it f, in the second case). Rather, c would have occurred just as it did but would have failed to cause e. It is less of a departure from actuality to get rid of e by holding c fixed and giving up some or other of the laws and circumstances in virtue of which c could not have failed to cause e, rather than to hold those laws and circumstances fixed and get rid of e by going back and abolishing its cause c. (In the second case, it would of course be pointless

not to hold f fixed along with c.) The causal dependence of e on c is the same sort of irreversible counterfactual dependence that we have considered already.

To get rid of an actual event e with the least over-all departure from actuality, it will normally be best not to diverge at all from the actual course of events until just before the time of e. The longer we wait, the more we prolong the spatiotemporal region of perfect match between our actual world and the selected alternative. Why diverge sooner rather than later? Not to avoid violations of laws of nature. Under determinism *any* divergence, soon or late, requires some violation of the actual laws. If the laws were held sacred, there would be no way to get rid of e without changing all of the past; and nothing guarantees that the change could be kept negligible except in the recent past. That would mean that if the present were ever so slightly different, then all of the past would have been different—which is absurd. So the laws are not sacred. Violation of laws is a matter of degree. Until we get up to the time immediately before e is to occur, there is no general reason why a later divergence to avert e should need a more severe violation than an earlier one. Perhaps there are special reasons in special cases—but then these may be cases of backward causal dependence.

Preemption

Suppose that c_1 occurs and causes e; and that c_2 also occurs and does not cause e, but would have caused e if c_1 had been absent. Thus c_2 is a potential alternate cause of e, but is preempted by the actual cause c_1. We may say that c_1 and c_2 overdetermine e, but they do so asymmetrically. [12]

> [12] I shall not discuss symmetrical cases of overdetermination, in which two overdetermining factors have equal claim to count as causes. For me these are useless as test cases because I lack firm naive opinions about them.

In virtue of what difference does c_1 but not c_2 cause e?

As far as causal dependence goes, there is no difference: e depends neither on c_1 nor on c_2. If either one had not occurred, the other would have sufficed to cause e. So the difference must be that, thanks to c_1, there is no causal chain from c_2 to e; whereas there is a causal chain of two or more steps from c_1 to e. Assume for simplicity that two steps are enough. Then e depends causally on some intermediate event d,

and d in turn depends on c_1. Causal dependence is here intransitive: c_1 causes e via d even though e would still have occurred without c_1.

So far, so good. It remains only to deal with the objection that e does *not* depend causally on d, because if d had been absent then c_1 would have been absent and c_2, no longer preempted, would have caused e. We may reply by denying the claim that if d had been absent then c_1 would have been absent. That is the very same sort of spurious reverse dependence of cause on effect that we have just rejected in simpler cases. I rather claim that if d had been absent, c_1 would somehow have failed to cause d. But c_1 would still have been there to interfere with c_2, so e would not have occurred.

Postscripts to "Causation"

A. Piecemeal Causation

Suppose that c and e are large, prolonged processes, each composed of many smaller events. Suppose it is not true (or not clearly true) that e, taken as a whole, causally depends on c, taken as a whole; suppose even that they are not connected by a chain of causal dependence. It may nevertheless be that c and e are divisible into parts in such a way that every part of e is causally dependent on (or connected by a chain of causal dependence to) some part of c. In that case we might well simply speak of c as a cause of e, though it is not so under the analysis I gave.

Self-sustaining processes exhibit piecemeal causation. For instance, suppose a public address system is turned up until it howls from feedback. The howling, from start to finish, is an event. If it had not occurred, it would not have occurred; but this is certainly not counterfactual dependence between *distinct* events, therefore it does not qualify as causal dependence on my account. Nor is there a closed causal loop, as in time travel stories, in which the howling causes itself because it depends causally on some distinct event which in turn depends causally on it. So it is not true, on my account, that the howling *taken as a whole* causes itself. What is true is that the howling

causes itself piecemeal. It is divisible into parts in such a way that each part except the first is caused by an earlier part, and each part except the last causes a later part. This causing of part by part is unproblematic: cause and effect are distinct events, wherefore their counterfactual dependence qualifies as causal. We might well say that the howling causes itself; this is to be accepted, but only in a derivative sense. Similarly, if two prolonged events sustain one another, each causes the other piecemeal. The example of the howling illustrates this case also: the sound in the air sustains the signal in the wires, and *vice versa.*

It may be that when we speak of causation in history we are often speaking of piecemeal causation. [1]

[1] Here I am indebted to a lecture given by Martin Putnam at Princeton in 1976.

A depression causes a wave of bankruptcies: what are we to make of this? If the depression had not occurred. . . . That is puzzling. To suppose away an entire depression takes us a long way from actuality. And the farther we depart from actuality, the more we lose control over our counterfactuals. For the more different respects of similarity and difference we have to balance, the more of a problem it is that we have left it vague just how to do the balancing, so the less clearly we know what is and what isn't to be held fixed in our counterfactualizing. (For instance, what if many of the firms that went broke came into existence during, and because of, the depression itself? Shall we hold their existence fixed in asking what would have happened without the depression?) But the depression is a big event that is divisible into many parts. Although it is hard to say what would happened without the entire depression, it is comparatively easy to say that without this or that event which was part of the depression, this or that one of the bankruptcies would not have taken place. Now, our counterfactuals are much more under control, because they stay much closer to home. So even if it is unclear what the depression taken as a whole might have caused, it is at any rate clear that various parts of it caused the various bankruptcies. That is to say that the depression was at least a piecemeal cause of the wave of bankruptcies.

There is a well-known dilemma about actions. Consider an action of raising my arm. First something goes on within my brain; then signals go out my nerves; then my muscles contract; and as they do, my arm rises. There seems to be a conflict between two things we want to say. (1) The action of raising my arm is a prolonged event with diverse parts. It is the whole causal process just described. It may begin within

end p.173

me, but it is not over until my arm rises. Its earlier parts cause its later parts, and its final part is the bodily movement. But (2) just as my action of raising a flag would be an event that causes a flag to rise, so my action of raising my arm is an event that causes my arm to rise. The raising, whether of flag or of arm, is so-called because it causes the rising.

Distinguish the inclusion of one event in another from mere involvement of one in the naming of the other. So far as involvement in naming goes, the two cases are on a par. For the flag and the arm alike, the raising deserves its name only if, and perhaps only after, a rising ensues. But with respect to inclusion, the two cases seem to differ. If I raise a flag by delayed action, I can be done raising it long before it rises. My action is over when I have done my part; the process that ends when the flag is up consists of more than just my action. (Beware ambiguity: the phrase "my raising the flag" might denote just my action, or it might denote the whole affair.) But if I raise my arm by delayed action—say that I have *very* sluggish nerves—then it takes me a long time to raise my arm. In this case, my part of the process is the whole of the process. So long as the signal is traveling through my sluggish nerves, so long as my muscles are contracting and my arm is rising, my part of the affair is still going on.

I would like to assent to both (1) and (2); the apparent obstacle is that we have two events, the raising and the rising, and according to (1) they are not wholly distinct, yet according to (2) one causes the other. But if this is a case of piecemeal causation, we have no problem. If an early part of the raising causes the rising which is a late part of the raising, we may still say simply that the raising causes the rising; just as, when an early part of the depression causes a bankruptcy which is a later part of the depression, we may still say simply that the depression causes the bankruptcy.

There is a second version of the problem. The rising of my arm is not the only event which is caused by the initial inner part of my action and yet takes place before my arm has risen. The same may be true of various side effects, events which definitely are not to be included as parts of the action. Suppose, for instance, that the nerves leading into my arm are monitored so that whenever I raise my arm the nerve signal produces a trace on an oscillograph. Because I can produce the trace by raising my arm, we ought to be free to say that my action causes the trace. And yet the trace appears before the arm rises. Shall we say that the effect precedes its cause? Or that the action which causes the effect is over sooner than we think? Neither: it is a case of piecemeal

end p.174

causation. Like the rising of the arm, the trace on the oscillograph is caused by an initial part of the action, and thereby is caused by the

action. [2]

[2] See Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980), II; and for the second version, see also G. H. von Wright, *Explanation and Understanding* (London: Routledge and Kegan Paul, 1971), pp. 76–81. I am indebted to Hornsby, and to Alison McIntyre, for discussion on this point.

B. Chancy Causation

In the paper, I confined my discussion to the deterministic case for the sake of brevity. [3]

[3] The paper was shortened at the request of the Program Chairman of the American Philosophical Association (Eastern Division). The full-length version (May 1973) advocated the same treatment of probabilistic causation that is presented in this postscript.

But I certainly do not think that causation requires determinism. (Hence I regard "causality" as a naughty word, since it is ambiguous between "causation" and "determinism.") Events that happen by chance may nevertheless be caused. Indeed, it seems likely that most actual causation is of just this sort. Whether that is so or not, plenty of people do think that our world is chancy; and chancy enough so that most things that happen had some chance, immediately before-hand, of not happening. These people are seldom observed to deny commonplace causal statements, except perhaps when they philosophize. An analysis that imputes widespread error is *prima facie* implausible. Moreover, it is dishonest to accept it, if you yourself persist in the "error" when you leave the philosophy room. We had better provide for causation under indeterminism, causation of events for which prior conditions were not lawfully sufficient.

One kind of chancy causation is already covered by my analysis, with no modification needed: c occurs, e has some chance of occurring, as it happens e does occur; but if c had not occurred, then e would have had no chance at all of occurring, and so would not have occurred. Then e depends causally on c, and c is a cause of e, according to my original analysis. So far, so good.

(Some would object to my step from "e would have had no chance of occurring" to "e would not have occurred." They say that things

with no chance at all of occurring, that is with probability zero, do nevertheless happen; for instance when a fair spinner stops at one angle instead of another, yet any precise angle has probability zero. I think these people are making a rounding error: they fail to distinguish zero chance from infinitesimal chance. Zero chance is *no* chance, and nothing with zero chance ever happens. The spinner's chance of stopping exactly where it did was not zero; it was infinitesimal, and infinitesimal chance is still *some* chance.)

But there is a second case to be considered: c occurs, e has some chance x of occurring, and as it happens e does occur; if c had not occurred, e would still have had some chance y of occurring, but only a very slight chance since y would have been very much less than x. We cannot quite say that without the cause, the effect would not have occurred; but we can say that without the cause, the effect would have been very much less probable than it actually was. In this case also, I think we should say that e depends causally on c, and that c is a cause of e.

It does not matter whether x itself, the actual chance of the effect, is high or low. Suppose you mischievously hook up a bomb to a randomizer—a genuinely chancy one, if need be one that works by counting clicks in a counter near a radioactive source. If you set the randomizer to a high probability, that makes it likely that your act of setting up the bomb will cause an explosion. If you set the randomizer to a low probability, that makes it less likely that your act will cause an explosion. But no matter how you set the randomizer, if the bomb does chance to go off, then your act does cause the explosion. For no matter how you set the randomizer, we can be sure that the explosion would have been very much less probable still if you hadn't set up the bomb at all.

(You took it in stride when you read my words: if you set the randomizer low, that makes it less likely that your act will cause an explosion. That proves my point. For suppose that improbable events cannot be caused: the actual chance x has to be high, or at least has to exceed some lowish threshold, in order to have a case of causation. Then if you set the randomizer low enough, that doesn't just make it *unlikely* that your act will cause an explosion—it makes it downright *impossible*. But "unlikely" did *seem* the right word. "Don't worry—set the randomizer below 0.17% and you can't *possibly* cause an explosion."—Not so!)

Several points of clarification may be helpful. (1) Chances are time-dependent: an event may have different chances at different times

before it occurs. The actual chance x of e is to be its chance at the time immediately after c; and the counterfactual is to concern chance at that same time. (2) I do not assume that there is some y that would definitely have been the chance of e in the absence of c. Maybe so, maybe not. Maybe in that case the chance of e might have had any of various values, all of them much less than the actual chance x. In saying that without c, e would have had some chance y much less than x, "some chance y" is a quantifier whose scope is limited to the consequent of the conditional. (3) "Much less" means less by a large factor—not by a large difference. If x is already small, the

difference of y and x could not be large. It is x that sets the standard for how small the chance of e must be without c. We could have one case in which the absence of a cause would lower the chance of an effect from 100% to 10%; another in which the lowering would be from 10% to 1%; yet another in which the lowering would be from 1% to 0.1%; . . . and all would count equally as cases of chancy causal dependence. So it will not do to simplify our counterfactual and say that without c, the chance of e would be low *simpliciter*. (4) A chance event may be caused; but we should not say that it is caused to happen *rather than not* . Contrastive causal statements differ from plain ones. According to what I say about contrastive questions and statements in "Causal Explanation" (in this volume), there can be no contrastive causal explanation of why a chance event occurs rather than not.

Many probabilistic theories of causation share the motivating idea that a cause increases the probability of the effect. Mine differs from some of the others in two respects. [4]

[4] For the other sort of probabilistic theories of causation, see *inter alia* Patrick Suppes, *A Probabilistic Theory of Causality* (Amsterdam: North-Holland, 1970); and Nancy Cartwright, "Causal Laws and Effective Strategies," *Noûs* 13 (1979): 419–37, reprinted with additions in her *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983). Cartwright does not offer her theory as an analysis; as such it would be circular, but it might nevertheless succeed as a constraint relating causation to probabilities.

An analysis much closer to mine, except that it does not provide for (what I would call) causation without causal dependence and it avoids reference to events, is that of D. H. Mellor, "Fixed Past, Unfixed Future," in Barry Taylor, ed., *Contributions to Philosophy: Michael Dummett* (The Hague: Nijhoff, 1986).

First, it is meant to apply to causation in the single case: causation by one particular event of another event, not conduciveness of one kind of event to another kind. Hence its probabilities are single-case chances, as opposed to finite or

limiting frequencies. You may not like single-case chances—I don't either—but I cannot see how to make sense of certain well-established scientific theories without them. If we need them anyway, we may as well use them here. (I discuss single-case chances, and the reason for disliking them, elsewhere in this volume; see "A Subjectivist's Guide to Objective Chance," especially the final section and Postscript C; also my discussion of Humean supervenience in the introduction.)

Second, my analysis is in terms of counterfactual conditionals about probability; not in terms of conditional probabilities. If we try to use an inequality of conditional probabilities to express that event c raises the probability of event e, we run into a well-known difficulty. The inequality may well hold not because c causes e, but rather because c and e are two effects of a common cause. One cure is to use fancier conditional probabilities: conditionalize not just on the absence of c, but on that together with a specification of background. Then the problem is to say, preferably without circular mention of causation, what information should be included in this background.

But even if that problem can be solved, another remains. Conditional probabilities, as standardly understood, are quotients. They go undefined if the denominator is zero. If we want to say, using conditional probabilities, that c raises the probability of e, we will need probabilities conditional on the non-occurrence of c (plus background, perhaps). But there is no guarantee that this conditional probability will be defined. What if the probability that c occurs (given background) is one? What if c has been predetermined through all of past time —what if its probability has *always* been one, so that even by going back in time we cannot find a non-zero chance of c's failing to occur? For that matter, what if we want to apply our probabilistic analysis of causation to a deterministic world in which all probabilities (at all times) are extreme: one for all events that do occur, zero for all that don't? The requisite conditional probabilities will go undefined, and the theory will fall silent. That is not acceptable. Earlier, I said that it would not do to impute error to indeterminists who accept commonplace causal statements; therefore we cannot accept an analysis of causation that works only under determinism. Likewise it would not do to impute error to determinists who accept commonplace causal statements; therefore we cannot accept an analysis that works only under indeterminism. An adequate analysis must be neutral. It must work in both cases. And it must work in a uniform way, for it does not seem that our concept of causation is disjunctive. A probabilistic analysis (of single-case causation) that uses conditional probabilities is

not neutral. It is made for indeterminism. My analysis, on the other hand, can serve alike under indeterminism or determinism. [5]

[5] It would be possible to squander this advantage of the counterfactual analysis, of course. One could interpret the counterfactuals themselves in such a way that they make non-trivial sense only under indeterminism: take as accessible counterfactual situations only those courses of events that once had some non-zero probability of coming to pass. A probabilistic theory of counterfactuals along these lines would make it child's play to confute the determinist out of his own mouth—an advantage that might commend it to some philosophers, but to me seems a sufficient *reductio*.

My motivating idea is that causes make their effects more probable; but that is written into the analysis of causal dependence, not of causation itself. As in my original analysis, we have causation when we have a causal chain: one or more steps of causal dependence. The effect need not depend on the cause directly. When we have causation without direct causal dependence, as in some cases of

preemption, it is not necessarily true that the cause at the beginning of the chain raises the probability of the effect at the end. The cause might lower the probability of the effect, or might leave it unchanged. At each step in the chain, we have a cause raising the probability of its immediate effect. But since counterfactuals are not transitive, that does not settle whether there is raising over the entire chain.

Suppose we have two redundant systems to produce the same effect. One is much more reliable than the other—that is, much less subject to random failure part way along the causal chain. The reliable system is already started; left to itself, it will very probably produce the effect. But I do not leave it to itself. There is a switch that both turns off the reliable system and turns on the unreliable system, and I throw this switch. As luck would have it, the unreliable system works. The effect ensues, just as it would probably have done without my act. My act did not make the effect more probable, but rather less, since I put the unreliable system in place of the reliable one. Nevertheless, I did cause the effect. And the reason is plain if we consider some intermediate event in the causal chain that actually occurred, something that happened well after the reliable system was already turned off. That event was part of the working of the unreliable system, so it would not have occurred, or at least it would have been improbable, if I had not thrown the switch. But by the time of the intermediate event, the reliable system was already out of action. So without that event, the effect would not have occurred, or at least it would have been very improbable. (Here it is crucial that the counterfactual be governed by a similarity relation that does not conduce to backtracking; see "Counterfactual

Dependence and Time's Arrow" in this volume.) My act raised the probability of the intermediate event, and thereby caused it. And the intermediate event raised the probability of the effect, and thereby caused it. So my act caused a cause of the effect, and thereby caused it—despite lowering its probability. [6]

> [6] Compare Wesley C. Salmon's discussion of "explanations that do not increase weight" in Salmon *et al.*, *Statistical Explanation and Statistical Relevance* (Pittsburgh: University of Pittsburgh Press, 1971), pp. 62–65. For a more wholehearted adherence to the thesis that causes make their effects more probable, see Nancy Cartwright, "Causal Laws and Effective Strategies," and D. H. Mellor, "Fixed Past, Unfixed Future."

I have said that if distinct events *c* and *e* both occur, and if the actual chance of *e* (at a time *t* immediately after *c*) is sufficiently greater than the counterfactual chance of *e* without *c*, that implies outright that *c* is a cause of *e*. Some philosophers find this counterintuitive. They would correct me thus.

> No; if there would have been some residual chance of *e* even without *c*, then the raising of probability only makes it *probable* that in this case *c* is a cause of *e*. Suppose, for instance, that the actual chance of *e*, with *c*, was 88%; but that without *c*, there would still have been a 3% probability of *e*. Then most likely (probability 97%) this is a case in which *e* would not have happened without *c*; then *c* is indeed a cause of *e*. But this just might be (probability 3%) a case in which *e* would have happened anyway; then *c* is not a cause of *e*. We can't tell for sure which kind of case this is.

It is granted, *ex hypothesi*, that it would have been a matter of chance whether *e* occurred. Even so, the objection presupposes that the case must be of one kind or the other: either *e* definitely *would* have occurred without *c*, or it definitely would *not* have occurred. If that were so, then indeed it would be sensible to say that we have causation only in case *e* definitely would not have occurred without *c*. My original analysis would serve, the amendment suggested in this postscript would be unwise, and instead of having a plain case of probabilistic causation we would have a probable case of plain causation.

But I reject the presupposition that there are two different ways the world could be, giving us one definite counterfactual or the other. That presupposition is a metaphysical burden quite out of proportion to its intuitive appeal; what is more, its intuitive appeal can be explained away.

The presupposition is that there is some hidden feature which may or may not be present in our actual world, and which if present would

make true the counterfactual that *e* would have occurred anyway without *c*. If this counterfactual works as others do, then the only way this hidden feature could make the counterfactual true is by carrying over to the counterfactual situation and there being part of a set of conditions jointly sufficient for *e*.

What sort of set of conditions? We think at once that the set might consist in part of laws of nature, and in part of matters of historical fact prior to the time *t*, which would together predetermine *e*. But *e* cannot be predetermined in the counterfactual situation. For it is supposed to be a matter of chance, in the counterfactual situation as in actuality, whether *e* occurs. That is stipulated as a hypothesis of the case. When an event is predetermined, there cannot be any genuine chance that it will not happen. Genuine chance gives us the residue of uncertainty that is left after *all* laws and prior conditions have been taken into account.

(Here I assume that we are not dealing with an extraordinary situation, involving time travel perhaps, in which the normal asymmetries of time break down, and the past contains news from the future. That is fair. The objection concerns what should be said about *ordinary* cases of probabilistic causation.)

So the hidden feature must be something else. But what else can it be? Not the historical facts prior to t, not the chances, not the laws of nature or the history-to-chance conditionals that say how those chances depend on the prior historical facts. For all those are already taken account of, and they suffice only for a chance and not a certainty of e.

There is the rest of history: everything that happens after t. These future historical facts are not relevant to the chances at t; e can still have a chance of not occurring even if there are facts of later history that suffice for its occurrence. As there will be: if it does occur, that is itself a fact of later history. In the terminology of "A Subjectivist's Guide to Objective Chance" (in this volume) later history is "inadmissible." So perhaps that is where the hidden feature of the world is to be found.

But this also will not do. For we know very well that if we give weight to future similarities, so that facts of the later history of our world tend to carry over into counterfactual situations, then we will get into trouble. We will get counterfactuals that seem false in themselves, and that also yield false conclusions about causation. We must make sure, either by fiat or else by tailoring our standards of similarity to exploit the *de facto* asymmetries of time, that future similarities will

end p.181

normally carry no weight. (See "Counterfactual Dependence and Time's Arrow" in this volume.) Features of our actual future history may be well hidden, sure enough, and they might well enter into sets of conditions and laws sufficient to postdetermine e, but what they will not do is carry over into the counterfactual situation without c.

(Normally. I am forced to admit exceptions of two kinds, for reasons discussed in Postscript D to "Counterfactual Dependence and Time's Arrow" in this volume. If a reconvergence to actual history could be accomplished without widespread miracles or quasi-miraculous coincidences, then I would admit that actual future history carries over into the counterfactual situation; and I would admit that the absence of such quasi-miracles carries over. But I think the first cannot apply to the truth of counterfactuals at a world like ours, and the second could apply only to the special case where e itself would be quasi-miraculous. So these exceptions are not relevant to our present discussion.)

So the hidden feature must be something else still: not a feature of the history of this world, and also not a feature of its chances, or of the laws or conditionals whereby its chances depend on its history. It fails to supervene on those features of the world on which, so far as we know, all else supervenes. To accept any such mysterious extra feature of the world is a serious matter. We need some reason much more weighty than the isolated intuition on which my opponent relies. [7]

[7] Some people *do* have more weighty reasons, though I do not think they are reasons that we ought to accept. Theological reasons, perhaps: if God is to be properly omniscient, and if He is to exercise divine providence without running risks, He had better know just what would happen if He made creatures whose choices were not predetermined. Then there have to be definite counterfactual facts for Him to know, even if they cannot supervene on any features of the world that we would otherwise believe in, and accordingly de Molina, Suarez, and (sometimes) Plantinga posited that there are these facts. See Robert M. Adams, "Middle Knowledge and the Problem of Evil," *American Philosophical Quarterly* 14 (1977): 109–17. Or physical reasons, perhaps: P. H. Eberhard, "Bell's Theorem without Hidden Variables," *Il Nuovo Cimento* 38 B (1977): 75–79, and likewise Nick Herbert and Jack Karush "Generalization of Bell's Theorem," *Foundations of Physics* 8 (1978): 313–17, fulfill the promise of their titles by appeal to a principle of "counterfactual definiteness". This principle says that even if a measurement was not made, and its outcome would have been a matter of chance if it had been made, nevertheless there is some definite value that it would have given. These counterfactual measurement outcomes do not supervene on the wave function which is the usual complete quantum mechanical description of a physical system. It is considered nice that we can get Bell's Theorem using just the counterfactual outcomes, instead of trafficking in hidden variables as traditionally conceived; though for my own part, I cannot tell the difference.

Without such a reason, it would be better to suppress the intuition.

end p.182

Which is all the easier if it rests on a mistake in the first place; and I think it does. I suspect that my opponent is someone who has not wholeheartedly accepted my stipulation of the case in question. Stipulation or no, he remains at least somewhat inclined to think that the case involves not genuine chance, but a kind of counterfeit chance that is compatible with determinism. (See Postscript B to "A Subjectivist's Guide to Objective Chance" in this volume.) Perhaps he clear-headedly thinks that counterfeit chance is all the chance there could ever be, and so is all that could be meant by the word "chance." Or perhaps he thinks double, and thinks of the case half one way and half the other.

If it *is* a case of counterfeit chance, then his objection is well taken. For then e is after all predetermined one way or the other, both in

actuality and in the counterfactual situation without *c*; but predetermined partly by details of prior historical fact that are far too minute to be discovered in advance. So we do indeed have an unproblematic hidden feature of the actual world—namely, the relevant configuration of minute details—that carries over to the counterfactual situation and there joins in predetermining the outcome one way or the other.

That is all very well, but then his objection is off target. I was not speaking of a case of counterfeit chance, I insist, but of a different case: probabilistic causation of a genuine chance event. If my opponent believes that my case is impossible because counterfeit chance is all the chance there can be, let him say so; but let him not reinterpret my case to fit his own doctrines.

When my opponent says that either *e* would have occurred without *c* or else *e* would not have occurred without *c*, he sounds like Robert Stalnaker. [8]

[8] See Robert C. Stalnaker, "A Theory of Conditionals," in Nicholas Rescher, ed., *Studies in Logical Theory* (Oxford: Blackwell, 1968); and "A Defense of Conditional Excluded Middle" in *Ifs*, ed. by William Harper *et al.* (Dordrecht: Reidel, 1980). I discuss Stalnaker's theory in "Counterfactuals and Comparative Possibility" and "Causal Decision Theory," both in this volume.

But his position is not the same, though he accepts the same disjunction of counterfactuals, and Stalnaker's defense of such disjunctions is no use to him. My opponent thinks there are two relevant ways the world might be; one of them would make true one of the disjoined counterfactuals, the other would make true the other, so the disjunction is true either way. Stalnaker, like me, thinks there is only one relevant way for the world to be, and it does not make either counterfactual determinately true. But Stalnaker, unlike me, thinks the disjoined counterfactuals are true or false relative to alternative arbitrary resolutions of a semantic indeterminacy; what makes the counterfactuals

lack determinate truth is that different resolutions go different ways; but every resolution makes one or the other true, so the disjunction is determinately true despite the complementary indeterminacies of its disjuncts. A resolution of an alleged semantic indeterminacy is not a hidden fact about the world; and that is the difference between Stalnaker and my opponent. Stalnaker disagrees with me on a small point of semantics; my opponent, on a large point of ontology. A resolution of an indeterminacy might indeed be *mistaken* for a hidden fact about the world—Stalnaker suggests, plausibly, that such mistakes are common. So if we accepted Stalnaker's view on the point of semantics, that would give us a second way to explain away my opponent's problematic intuition.

C. Insensitive Causation

Killing, so they say, is causing to die. I am sure that I—and likewise you, and each of us—have caused ever so many people to die, most of them people yet unborn. Acts of mine are connected to their deaths by long chains of causal dependence. [9]

[9] Not acts of omission, if such there be. In the next postscript I shall consider causation by omission, but for the present I am discussing cases in which we have what is uncontroversially a genuine act—or more generally, a genuine event—to do the causing.

But I have never killed anyone—I hope.

For instance, suppose I write a strong recommendation that lands someone a job; so someone else misses out on that job and takes another; which displaces a third job-seeker; this third job-seeker goes elsewhere, and there meets and marries someone; their offspring and all their descendants forevermore would never have lived at all, and *a fortiori* would never have died, and so presumably their deaths would not have occurred, but for my act. [10]

[10] It has been observed in other connections that who will live in the future is a very sensitive matter, depending very much on the great and small events of the present and past. See Derek Parfit, "Future Generations: Further Problems," *Philosophy and Public Affairs* 11 (1982): 113–72; and Robert M. Adams, "Existence, Self-Interest, and the Problem of Evil," *Noûs* 13 (1979): 53–65.

Maybe there is a time after which *every* death that occurs is one that would not have occurred but for my act. It would be strange to single out my act as *the* cause of all those deaths. But it is *a* cause of them, under my analysis and also according to our common usage. And still I deny that I have ever killed.

For a still more striking case, consider the Big Bang. This event, I

take it, is a cause of *every* later event without exception. Then it is a cause of every death. But the Big Bang did not kill anyone.

So killing must be a special kind of causing to die. But what distinguishes this special kind of causation?

Not that there must be one single step of causal dependence, as opposed to an intransitive chain. An act of killing can be a preempting

cause. It can be you who kills the victim, even though another killer was standing by who would have done the job for you—causing the victim to die the very same death—if he had not seen you lay the poison yourself.

Not that the chance of the effect must be high. If you hook up a bomb to a randomizer and hide it in a crowded place, and it happens to go off, you can kill no matter how low you set the chance.

Not that the causal chain must be short. You can kill by delayed action. If you set a hidden time bomb with a thousand-year fuse, you may well kill someone yet unborn.

Not that the chain must be simple. You can kill someone by means of a lethal Rube Goldberg machine.

Not that the chain must be foreseeable. You can kill someone no matter how good your reasons were for thinking the gun was not loaded; or no matter how unfeasible it would have been for you to discover in advance his lethal allergy to what you fed him.

Not that the chain must pass through no later human actions. [11]

> [11] *Pace* Jennifer Hornsby, *Actions*, pp. 127–30. While disagreeing with Hornsby's general claim, I disagree less about the examples that motivate it; examples in which somebody causes a dinghy to sink by ordering someone else to sink it, or causes a death by ordering someone else to kill. See the final part of this postscript. I am indebted on this point to discussion with Hornsby.

If you kill by setting a baited mantrap, or by making a gift of poisoned chocolates, your unsuspecting victim's action is an intermediate step in the causal chain whereby you kill him. In other cases, an action by a third party may be an intermediate step: you make a gift of poisoned chocolates to the host, who offers them to the guest.

Perhaps a cluster of these conditions, inadequate if taken one by one, would work to distinguish the kind of causing that can be killing. I think not. But the counterexamples get too contrived to be very persuasive: imagine a lethal Rube Goldberg machine with a randomizer at one step, a thousand-year fuse at another, an alternative waiting in reserve at another, dependence on some action of the unsuspecting victim at another, and no way to discover how it works.

end p.185

I suggest a different way to distinguish the right kind of causing: by its insensitivity to circumstances. When an effect depends counterfactually on a cause, in general it will depend on much else as well. If the cause had occurred but other circumstances had been different, the effect would not have occurred. To the extent that this is so, the dependence is sensitive. Likewise if a causal chain consists of several steps of causal dependence, we can say that the chain is sensitive to the extent that its steps are. (On average? Or at worst?) Sensitivity is a matter of degree, however. It may be that the causation depends on an exceptionally large and miscellaneous bundle of circumstances all being just right. If any little thing had been different, that cause would not have caused that effect. But sometimes causation is comparatively insensitive to small differences in the circumstances. When my strong recommendation causes lives and then deaths, that is comparatively sensitive causation—there are many differences that would have deflected the chain of events. But if you shoot at your victim point-blank, only some very remarkable difference in circumstances would prevent his death. The same is true if you set a Rube Goldberg machine, or a delayed-action bomb, working inexorably toward its lethal outcome. The case of the bomb with a randomizer also is comparatively insensitive: the bomb might very well have chanced not to go off, but it isn't the fine details of the circumstances that would make the difference.

Jonathan Bennett restates my suggestion this way: killing requires "that the causal chain run through a stable and durable structure rather than depending on intervening coincidental events." [12]

> [12] "Killing and Letting Die," in Sterling W. McMurrin, ed., *The Tanner Lectures on Human Values*, Volume II (Cambridge: Cambridge University Press, 1981), p. 71.

A lethal Rube Goldberg machine may work in many steps, it may be full of thousand-year fuses and randomizers and alternatives waiting in reserve, its working may require the responses of unsuspecting agents, there may be no way to discover how it is built or understand how it would work; and yet it may be no less "stable and durable" for all that, and the causal chain running through it may be far more independent of "intervening coincidental events" than are most of the causal chains in the wider world.

So it seems that the reason why a lot of causing to die is not killing is, at least partly, that the causing to die in killing must be causation of a comparatively insensitive kind. And if this is so for killing, perhaps it is so likewise for other causatives. Consider the ways in which you can and can't make, break, wake, or bake things.

end p.186

Insensitivity is not the same thing as any of the unsatisfactory conditions that I considered above; but of course it is connected to several of them. *Ceteris paribus*, shortness and simplicity of the chain will make for insensitivity; insensitivity, in turn, will make for foreseeability.

The more the chain depends on a lot of circumstances being just right, the harder it is for a would-be predictor to know all he needs to know about the circumstances. The sensitivity of the chain is an obstacle to prediction. Unforeseeability does not imply sensitivity, since any of many other obstacles to prediction might be at work. But unforseeability sets a minimum standard. If a chain is insensitive enough that you can predict it, then it is insensitive enough that you can kill by it. Perhaps our common knowledge of what can normally be predicted sets a common standard for everyone. Or perhaps the standard varies. What if you are much better than I am at predicting chains that are somewhat sensitive? I am inclined to say that if so, then indeed you can kill in ways that I cannot. If your act and mine cause death by chains that are exactly alike, and if the duplicate chains are insensitive enough to fall within your powers of prediction but sensitive enough to frustrate mine, then you kill but I do not. [13]

[13] At this point, I am indebted to Jonathan Bennett.

My suggestion faces a problem. Recall that you can kill by a causal chain that has someone else's action as an intermediate step: you give someone poisoned chocolates, he unsuspectingly serves them to his guest, and thereby you kill the guest. (It is true as well that the host unwittingly kills the guest. But that is beside the point; the question is whether you kill the guest, and I submit that you and the host both do.) But if you tell the host that the chocolates are poisoned, and you order or hire or coerce or persuade him to serve them anyway, then it seems that you do not kill the guest. You may be no less guilty, morally and in the eyes of the law, than if you had killed him; or no less praiseworthy, if the guest was Hitler. Be that as it may, it seems that you don't kill by getting someone else to kill *knowingly*. Why not, on the suggestion I have advanced? It seems that if someone else is ready to kill knowingly when ordered or hired or coerced or persuaded to, his readiness well might be a stable and durable structure; so that by depending on this readiness, the causal chain from your action to a death well might be fairly insensitive to fortuitous circumstances.

I reply that indeed that might be so, and nevertheless we might speak as if it were not so. That would be no surprise. Part of our habitual

respect for other people consists in thinking that they are sensitive to a great variety of considerations, and therefore not easy to predict or control. It is all very well to take for granted that someone is ready to offer a guest what he takes to be harmless chocolates; to that extent, it is not disrespectful to regard his dispositions as a stable and durable structure. Offering chocolates is no big deal. It is another thing to take for granted that someone can be ordered or hired or coerced or persuaded to kill knowingly. That is to take him altogether too much for granted. The relevant disrespect lies not in thinking him willing to kill; whether that is disrespectful depends on the circumstances and the victim. Rather, it lies in thinking of his readiness to kill as stable and durable, inexorable, insensitive to fortuitous circumstances of the case, so that he is disposed to make weighty choices with unseemly ease.

Such disrespect might be well deserved. We might know very well that this dull thug before us would never think twice about killing for a small fee. Therefore, we might be sure that when you hire him, the causal chain from your action to the victim's death is as inexorable and insensitive as if it had passed instead through some strong and sturdy machine. But we might know this, and yet be halfhearted in putting our mouths where our minds are. Some vestige of our habitual respect might well influence how we speak. If I am right, when you cause death by hiring this thug, you are in literal truth a killer, no less than the thug himself is. If we deny it, I suggest that we are paying the thug a gesture of respect—insincere, undeserved, yet unsurprising.

That was an uncompromising version of my reply. I can offer an alternative version that runs as follows. If you hire the thug just considered, you are *not* in literal truth a killer. The truth conditions for "kill" are not just a matter of insensitive causation. They make an exception for insensitive causal chains that run through someone else's action of knowingly killing. However, insensitivity remains the underlying idea. The extra twist in the truth conditions is not just a brute complication of the concept; it is there, understandably, thanks to our respectful presumption that a causal chain through someone else's weighty decision will not be insensitive. The two versions agree about what we say, and why we say it; they differ only about what is literally true. *Ceteris paribus* it is bad to claim that we say what we know is literally false; but *ceteris paribus* it is bad to build complicating exceptions into the conditions of literal truth. Between the version that does one and the version that does the other, I think there is little to choose. I am not even confident that there is a genuine issue between the two.

D. Causation by Omission [14]

[14] In this postscript, I am much indebted to discussion with Jonathan Bennett and with Alison McIntyre.

An omission consists of the nonoccurrence of any event of a certain sort. To suppose away the omission is, exactly, to suppose that some event of the given sort does occur. We say that omissions may be caused, and may cause, and I have no wish to deny this. I would like to be able to provide for causation by omission within the general framework given in this paper and in "Events" (in this volume).

Unfortunately, I do not see how to make it fit with all that I say in general about events and about their causal dependence. So, one way or another, a special case it must be.

Omissions as effects are no special problem. I must allow in any case that sometimes, by causing suitable events, causes can create a pattern of events; and that a fact can supervene on this pattern even when there is no genuine event that can be called the obtaining of that fact; in which case the causes of the events in the pattern can also be said to cause that fact to obtain. [15]

[15] Here I do not rely on any fancy theory of facts; they are simply truths. That is to say they are the true ones among whatever entities can be said to bear truth values. On this view, as opposed to some fancy theories, most facts are only accidentally facts. They are contingent truths, and might have been falsehoods.

For instance, it is at least a fact that Xanthippe became a widow. I think there is no genuine event that can be called Xanthippe's becoming a widow. But the causes of her marriage together with the causes of Socrates's death may nevertheless be said to have caused her to become a widow: they caused genuine events that comprised a pattern on which the fact that she became a widow supervened. Certainly this fact is not beyond the reach of causal explanation. Likewise I can say that various distractions caused Fred to omit the precautions he should have taken; and in saying this, I needn't grant that there was any such thing as an event of omission. If there are events of omission, well and good. But I don't need them as effects.

Do I need them as causes? There are two opposite strategies that I might follow, and a third which is a compromise between those two. One way or another, all of them treat causation by omission as a special case. While I would guess that any of the three could be made to work, I am not in a position to prove it by presenting fully developed versions. I am not sure how much the three really differ; certainly some of their difference is just terminological.

end p.189

The first strategy accepts that there are events of omission. What is more, there are events essentially specifiable as omissions. For instance, Fred's omission of precautions, essentially specifiable as such, is an event that would have occurred no matter how he omitted them, no matter what else he did instead; and that could not have occurred if he had taken the precautions. For any event, there are necessary and sufficient conditions, normally hard to state, for that very event to occur. Some descriptions of an event are built into its conditions of occurrence; others are not. The first strategy says that the description of this event as an omission *is* built in. Then to suppose counterfactually that this event of omission does not occur is equivalent to supposing that Fred does take the precautions. So the counterfactual analysis of causation can apply to events of omission just as it does to all other events; and it is safe to say, as we ought to, that the effects which depend causally on Fred's omission are those which would not have occurred if he had taken the precautions. This strategy requires no exception to what I say about causation in general.

But it does require an exception to what I say about events in general. For I say that a theory of events, if it is built to serve the needs of my analysis of causation, must reject overly disjunctive events. An alleged event would be disjunctive if, or to the extent that, it could have occurred in various dissimilar ways. (The point is not that its conditions of occurrence could be formulated as a disjunction—anything can be formulated as a disjunction—but that they could be formulated as a disjunction of overly varied disjuncts.) An alleged event that is essentially specifiable as a talking-or-walking, and which could have occurred either as a talking or as a walking, is an example of what ought to be rejected. The reason is that if it were accepted as an event, then it could qualify as a cause; but it is intuitively very wrong to say that the talking-or-walking causes anything. But if we are to accept events of omission, in the way we are considering, then we may not reject disjunctive events without exception. For an event of omission, essentially specifiable as such, is highly disjunctive. Fred omits the precautions if he does something else during the period in which he was supposed to attend to them. So there are as many different ways for the event of omission to occur as there are alternative ways for Fred to spend the time. An event essentially specifiable as an omission amounts to an event essentially specifiable as a sleeping-or-loafing-or-chatting-or- . . . with a disjunct for everything Fred might do other than attending to the precautions. If omissions are accepted as genuine events and as causes, while other alleged disjunctive events are rejected,

end p.190

that makes causation by omission a special case. [16]

[16] Compare Jonathan Bennett's account of the distinction between killing and letting die in " 'Whatever the Consequences'," *Analysis* 26 (1966): 83–102, especially pp. 94–96. He presents the same distinction more fully in "Killing and Letting Die"; but there gives it a new name—positive versus negative instrumentality—because he observes that other considerations somewhat affect the ordinary usage of the ordinary terms. I agree, but shall ignore those considerations here.
There are ever so many ways you might move (or hold still—let us count this as one "way of moving") during a period of time. Suppose that if you were to move in any way that falls within the range *L*, someone would live; whereas if you were to move in any way that falls within the complementary range *D*, he

would die; and you move in a way that falls within range D, so he dies. Have you killed him? Or have you let him die, in other words omitted to save his life? (To avoid irrelevant issues, suppose (1) that the dependences are insensitive in the sense of the previous postscript; (2) that he would die the same death no matter how you moved within the range *D*; and (3) that this is not one of those special cases in which you could be said *both* to kill someone *and* to let him die, and by the very same conduct.) Bennett suggests, I think rightly, that if the range *L* is wide and varied compared to *D*, then you have killed him; whereas if the range *L* is narrow and uniform compared to *D*, then you have let him die.

I note that if the range *D* is wide, then an alleged event essentially specified as a moving-some-way-in-*D* is disjunctive and therefore suspect; and this suspect event would be essentially specified as a letting-die, and thus as an event of omission. Not so if the range *D* is narrow. On the strategy presently under consideration, there are such events of omission; on the strategy to be considered next, there are not. Of course I am not suggesting that these two strategies have different moral implications. Whatever events there may or may not be, what matters is that someone's life depended on how you moved.

The unfinished business for this strategy, of course, is to draw the line: how do we distinguish the genuine omissions from other alleged events that we should still reject? For instance, I think we ought not to say that the laws of nature, or other regularities, cause things; yet regularities may be made out to be omissions on a cosmic scale—the universe omits to contain events that would violate them. We must somehow deny that we have here a genuine event of omission.

The second, opposite, strategy says that there are no events of omission. Then there is no need to make a place for them within a theory of events, and no need to worry that they would be like other alleged events that are to be rejected. So far, so good. But in that case, I need to make an exception to what I say about causation itself. For it is not to be denied that there is causation by omission; and I cannot analyze this in my usual way, in terms of counterfactual dependence between distinct events. Instead I have to switch to a different kind of counterfactual for the special case. The counterfactual is not: if event *c* (the

omission) had *not* occurred. . . . It is rather: if some event of kind *K* (the omitted kind) *had* occurred. . . .

But if we use special counterfactuals for the special case, that opens several questions. Again we need to draw a line. I thought it necessary to block some counterexamples against a counterfactual analysis of causation by insisting that counterfactual dependence was to be between events. If we give that up, what new line shall we retreat to? As before, alleged causation by the laws of nature, regarded as cosmic omissions, will illustrate the problem. Also, I thought it necessary to insist on distinctness between events that stand in causal dependence, and by distinctness I meant more than nonidentity. (See "Events.") But how does distinctness apply to causation by omission? Fred sleeps, thereby omitting precautions against fire and also omitting precautions against burglary. Two distinct omissions?

The third, compromise strategy accepts events of omission as causes; but this time, the events of omission are not essentially specified as such. Fred omits the precautions, sleeping through the time when he was supposed to attend to them. His nap was a genuine event; it is not objectionably disjunctive. There are many and varied ways in which he could have omitted the precautions, but there is just one way that he did omit them. We could plausibly say, then, that his nap *was* his omission of precautions. But accidentally so. His nap could have occurred without being an omission of precautions: if (1) that very nap had been taken somewhat later, with the precautions seen to beforehand; or conceivably (2) if he had taken the precautions somehow in his sleep, or (3) if that very nap could have been taken by someone else, or (4) if the precautions had not been his responsibility. (I take it that (2)–(4) are problematic in various ways; so I rest my case mainly on (1).) And an omission of precautions might very well have occurred without being that nap: he might have stayed awake and done any of many other things instead of attending to the precautions. Still, *as it was*, the nap was what happened instead of the taking of precautions. So we may call it an event of omission, though we do not thereby capture its essence. We can have events of omission, so understood, and still reject disjunctive events without exception.

But this third strategy, like the second, demands special counterfactuals for the special case. Even if Fred's nap was his omission of precautions, it is one thing to suppose that this very event did not occur, and it is another thing to suppose that no event that occurred (this or any other) was an omission of the precautions. It is one thing to suppose away the event *simpliciter*, another thing to suppose it away *qua*

omission. It is the second counterfactual supposition, not the first, that is relevant to causation by omission. For it is the second supposition that is equivalent to supposing that the precautions were taken. But this is special. In other cases the relevant counterfactuals are those that suppose away an event *simpliciter*; we do not in general need to suppose away events *qua* satisfying some or other accidental specification.

As with the second strategy, resort to special counterfactuals for the special case threatens to undo our defenses against various counter-examples. It remains to be seen how, if at all, those defenses could be rebuilt. This strategy, like the others, leaves us with unfinished business.

E. Redundant Causation [17]

[17] In this postscript I am much indebted to discussion with John Bigelow, with John Etchemendy, and with Louis Loeb.

Suppose we have two events c_1 and c_2, and another event e distinct from both of them; and in actuality all three occur; and if either one of c_1 and c_2 had occurred without the other, then also e would have occurred; but if neither c_1 nor c_2 had occurred, then e would not have occurred. Then I shall say that c_1 and c_2 are *redundant causes* of e.

(There might be redundant causation with a set of more than two redundant causes. There might be probabilistic redundant causation, in which e would have had some small chance of occurring even if neither c_1 nor c_2 had occurred. There might be stepwise redundant causation without direct dependence, as described by Louis Loeb. [18]

[18] "Causal Theories and Causal Overdetermination," *Journal of Philosophy* 71 (1974); 525–44 ▭ Link ▸ . The simplest stepwise case is as follows; there could be more steps, or more events at any step. We have five actual events: c_1, c_2, d_1, d_2, e, with e distinct from the d's and the d's distinct from the c's. If neither of the c's had occurred, then neither of the d's would have occurred; but if either of the c's had occurred alone, then one of the d's would have occurred. If neither of the d's had occurred, then e would not have occurred; but if either of the d's had occurred alone, then e would have occurred. So the c's redundantly cause e by way of the d's. But if neither of the c's had occurred, e would have occurred anyway; so we do not have direct redundant causation of e by the c's.

I pass over these complications and consider redundant causation in its simplest form.)

As in my definition of ordinary causation, the counterfactuals concern particular events, not event-kinds. So it is not redundant causation if you shoot a terminal cancer patient—or, for that matter, a healthy young mortal—who would sooner or later have died anyway. Without your act

he would have died a different death: numerically different, because very different in time and manner. The particular event which is the death he actually dies would not have occurred. If you shoot a man who is being stalked by seven other gunmen, that *may* be redundant causation—the answer depends partly on details of the underdescribed case, partly on unsettled standards of how much difference it takes to make a different event. If you shoot a man who is simultaneously being shot by seven other members of your firing squad, that doubtless is redundant causation. The exact number of bullets through the heart matters little.

If one event is a redundant cause of another, then is it a cause *simpliciter*? Sometimes yes, it seems; sometimes no; and sometimes it is not clear one way or the other. When common sense delivers a firm and uncontroversial answer about a not-too-far-fetched case, theory had better agree. If an analysis of causation does not deliver the common-sense answer, that is bad trouble. But when common sense falls into indecision or controversy, or when it is reasonable to suspect that far-fetched cases are being judged by false analogy to commonplace ones, then theory may safely say what it likes. Such cases can be left as spoils to the victor, in D. M. Armstrong's phrase. We can reasonably accept as true whatever answer comes from the analysis that does best on the clearer cases. It would be still better, however, if theory itself went indecisive about the hard cases. If an analysis says that the answer for some hard case depends on underdescribed details, or on the resolution of some sort of vagueness, [19]

[19] Or on resolution of an ambiguity. Loeb (*op. cit.*) has offered a counterfactual analysis of causation in a broad sense—he calls such causes "C-conditions"—which would include redundant causes whether or not they are causes on my narrower analysis. Likewise Ardon Lyon's counterfactual analysis in "Causality," *British Journal for the Philosophy of Science* 18 (1967): 1–20, is modified so that it includes some redundant causes. I fear that such analyses, though perhaps suited to Loeb's purpose in formulating causal theories of memory *et al* ., are too broad to correspond to any ordinary sense of the word "cause". Be that as it may, it remains possible that the hard cases are causes in one sense but not in another. If so, then if the counterfactual approach is right, it ought to afford analyses for all the senses.

that would explain nicely why common sense comes out indecisive.

In my paper, I distinguished one kind of case—preemption with chains of dependence—in which common sense delivers clear positive and negative answers, and my counterfactual analysis succeeds in agreeing. I left all other cases of redundant causation as spoils to the victor, doubting that common-sense opinions about them would be firm and uncontroversial enough to afford useful tests of the analysis.

Now I would distinguish more varieties of redundant causation. Sometimes my analysis, as it stands, agrees with clear common-sense answers, positive or negative. Sometimes it reproduces common-sense indecision. Sometimes I am still content to leave far-fetched cases as spoils to the victor. But sometimes it seems that additions to my original analysis are needed.

I consider first a class of cases distinguished by doubt as to whether they exhibit redundant causation at all. I have already mentioned one example: you shoot a man who is being stalked by seven other gunmen. As it actually happens, the man dies on Tuesday morning, face down on the ground, his heart pierced by your bullet, with an entry wound in his back and an exit wound in his chest. Without your act he would have died on Wednesday evening, slumped in a chair, his heart pierced by someone else's bullet, with an entry wound in his chest and an exit wound in his back.

Is it that without your act he would have died a different death—numerically different because somewhat different in time and manner? [20]

[20] Here and in what follows, I assume that difference in time or manner is what it takes to make a numerical difference between an event that actually occurs and one that would have occurred under some counterfactual supposition. That is contrary to a view put forward by Peter van Inwagen in "Ability and Responsibility," *Philosophical Review* 87 (1978): 201–24, especially pp. 208–209, and in *An Essay on Free Will* (Oxford: Clarendon Press, 1983), pp. 167–70. He suggests that an event which actually occurs as the product of certain causes could not have occurred without being the product of those causes; nor could those causes have had a different event as their product. He finds this view plausible in part because of its analogy to the view that human beings, tables, etc. should be individuated by their causal origins.

I reject his view for two reasons. First, because it would ruin my project of analyzing causation in terms of counterfactual dependence. It would trivialize any counter-factual to the effect that without the cause, the effect would not have occurred. Second, because it is *prima facie* implausible. I can legitimately entertain alternative hypotheses about how an event (or for that matter a human being, or a table) was caused; or I can entertain alternative plans about how some desired future event is to be caused. But if I do, then I certainly seem to be presupposing that one and the same event might be produced by various different causes. (Compare van Inwagen's own remark that we seem to presuppose that one and the same event might have had different effects.) But van Inwagen's view implies that things are not as they seem: either my hypotheses or plans (with at most one exception) are hidden impossibilities; or else they are not about a particular event at all, but rather they involve some highly specific kind of event. These reconstruals seem artificial, and not to be accepted without better reason than van Inwagen gives.

If so, there is no redundancy. The particular death he actually died depends counterfactually on your act, without which that very

event would not have occurred. This is straightforward causation. Or is it rather that without your act he would have died the very same death—numerically the same, despite slight differences in time and manner? If so, there is genuine redundancy. In that case your act would be a redundant cause; whether it would be a cause *simpliciter* awaits our discussion of the varieties of genuine redundant causation.

It is hard to say which is true. It would remain hard, I think, no matter how fully we described the details of what actually happened, and of what would have happened under our counterfactual hypothesis.

Here is another example. Suppose three neurons are hooked up thus.
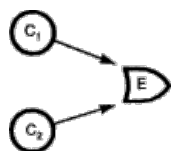
Suppose that a neuron fires if stimulated by the firing of one or more other neurons connected to it by a stimulatory synapse (shown by a forward arrowhead). But suppose—fictitiously, I believe—that a neuron fires much more vigorously if it is doubly stimulated than if it is singly stimulated. Neurons C_1 and C_2 fire simultaneously, thereby doubly stimulating E, which fires vigorously. Is this vigorous firing of E a different event from the feeble firing that would have occurred if either one of C_1 and C_2 had fired alone? Then we have joint causation, in which the effect depends counterfactually on each of the causes, and there is no redundancy. Or is it that numerically the same firing would have occurred, despite a difference in manner, with single stimulation? Then we have redundant causation. Again it is hard to say, and again the difficulty cannot be blamed on underdescription of the details.

Call an event *fragile* if, or to the extent that, it could not have occurred at a different time, or in a different manner. A fragile event has a rich essence; it has stringent conditions of occurrence. In both our examples we have redundant causation if the effect is not too fragile, ordinary causal dependence on joint causes otherwise.

Don't say: here we have *the events*—how fragile are *they*? Instead it should be: here we have various candidates, some more fragile and some less—which ones do we call the events? (For instance under my proposal in "Events," in this volume, the candidates will be smaller and larger classes of possible spatiotemporal regions, more and less

tightly unified by similarity.) Properly posed, the question need not have a fully determinate answer, settled once and for all. Our standards of fragility might be both vague and shifty.

As of course they are. You can say: the performance should have been postponed until the singer was over his laryngitis; then *it* would have been better. You can just as well say, and mean nothing different: the performance should have been cancelled, and another, which would have been better, scheduled later to replace it. There's no right answer to the question how fragile the performance is. Not because there is something—the performance—with an indeterminate size in logical space! But because there are various things, with various sizes, and we haven't troubled to decide which one is "the performance." Likewise every region of the earth has exact boundaries and a determinate size. Silicon Valley, whatever exactly that is, is no exception. However we haven't decided exactly how big a region is called "Silicon Valley." That's why there's no right answer to the question whether these words (written on the Stanford campus) were written in Silicon Valley.

So there may be no right answer to the question whether we have a case of joint causation without redundancy; or whether instead we have a case of redundant causation, which might or might not count as causation according to considerations to be discussed later. The answer depends on the resolution of vague standards of fragility. If common sense falls into indecision and controversy over such cases, that is only to be expected.

It is a common suggestion to adopt extreme standards of fragility, and thereby make away with redundant causation altogether. Even if a man is shot dead by a firing squad, presumably it would have made *some* minute difference to the time and manner of his death if there had been seven bullets instead of eight. So if you fired one of the eight bullets, that made some difference; so if his death is taken to be very fragile indeed, then it would not have occurred without your act. Under sufficiently extreme standards of fragility, the redundancy vanishes. Even this turns out to be a case in which the effect depends on each of several joint causes. Likewise for other stock examples of redundancy.

(Suppose we did follow this strategy wherever we could. Wouldn't we still have residual cases of redundancy, in which it makes *absolutely* no difference to the effect whether both of the redundant causes occur or only one? Maybe so; but probably those residual cases would be mere possibilities, far-fetched and contrary to the ways of this world. Then we could happily leave them as spoils to the victor. For we could plausibly suggest that common sense is misled: its habits of thought are

formed by a world where every little thing that happens spreads its little traces far and wide, and nothing that happens thereafter is quite the same as it would have been after a different past.)

Extreme standards of fragility would not fit a lot of our explicit talk about events. We do say—within limits!—that an event could have been postponed and could have happened differently. But this is not a decisive objection. The standards that apply within the analysis of causation might differ from those that apply in explicit talk.

What matters more is that extreme standards would not fit a lot of our negative judgements about causation itself. Extreme fragility of effects would make for spurious causal dependence in many quite ordinary cases. It would make more trouble than it cures. [21]

> [21] I owe this point to Ken Kress, *circa* 1968.

For instance, suppose there was a gentle soldier on the firing squad, and he did not shoot. If the minute difference made by eight bullets instead of seven is enough to make a different event, then so is the minute difference made by eight instead of nine. So if the victim's death is so very fragile that it would not have occurred without your act, equally it is so fragile that it would not have occurred without the gentle soldier's omission. If by reason of fragility the death depends causally on your act, then equally it depends causally on the omission. So the gentle soldier caused the death by *not* shooting, quite as much as you caused it by shooting! This is a *reductio*.

That case may puzzle us because it involves at least an appearance of redundancy, and also because it involves causation by omission. But the problem arises for cases without these complications. Boddie eats a big dinner, and then the poisoned chocolates. Poison taken on a full stomach passes more slowly into the blood, which slightly affects the time and manner of the death. If the death is extremely fragile, then one of its causes is the eating of the dinner. Not so.

To be sure, resolution of vagueness is influenced by context; and I can imagine a special context in which we might after all agree that the eating is a cause of the death. Pleased that Boddie is dead but horrified that the death was lingering, the poisoner says: if only he hadn't eaten, *this* wouldn't have happened—and by "this" he means the death, taken as very fragile. Maybe indeed that context makes it right to say that the eating caused the death. But it is also right, certainly in other contexts and probably even in this one, to say what is true under more lenient

and more ordinary standards of fragility: namely, that the eating did *not* cause the death. [22]

[22] How can it ever be right to say *A*, and equally right to say not-*A*?—Because sometimes what you say is itself the decisive part of the context that resolves vagueness and sets the standards whereby the truth value of what you say is determined. Say *A*, and thereby you set standards under which *A* is true, so you speak truly. But say not-*A* instead, and you speak just as truly; for in that case you set standards under which *A* is false. See "Scorekeeping in a Language Game" in my *Philosophical Papers* , Volume I.

So if we wanted to make away with the stock examples of redundant causation, what we would require is not a uniformly stringent standard of fragility, but rather a double standard—extremely stringent when we were trying to show that an effect really depends on its alleged redundant causes, but much more lenient when we were trying to agree with common-sense judgements that an effect is not caused by just anything that slightly affects its time and manner. It is not out of the question that there should be such a double standard. But if there is, an adequate theory of causation really ought to say how it works. (The changes of standard noted above, brought on by contextual pressures, are not the ones we want—they cut across cases with and without apparent redundancy.) To say how the double standard works may not be a hopeless project, but for the present it is not so much unfinished as unbegun.

Extreme fragility of effects might get rid of all but some far-fetched cases of redundant causation, but it leads to trouble that we don't know how to control. Moderate fragility gets rid of some cases and casts doubt on others, but plenty are left. Our topic has not disappeared.

So I return now to genuine redundant causation, including the doubtful cases when taken under standards of fragility that make them genuine. I divide it into *preemption* and (*symmetrical*) *overdetermination*. [23]

[23] I shall use the word "overdetermination" narrowly, to imply symmetry and exclude cases of causal preemption.

In a case of preemption, the redundant causes are not on a par. It seems clear that one of them, the *preempting cause*, does the causing; while the other, the *preempted alternative* , waits in reserve. The alternative is not a cause; though it could and would have been one, if it had not been preempted. There is the beginning of a causal process running from the preempted alternative to the effect. But this process does not go to completion. One effect of the preempting cause is to cut it off. In a case of overdetermination, on the other hand, there is no such asymmetry between the redundant causes. It may or may not be

clear whether either is a cause; but it is clear at least that their claims are equal. There is nothing to choose between them. Both or neither must count as causes.

First, preemption. It is clear what answer we want—the preempting cause is a cause, the preempted alternative is not—and any analysis that does not yield that answer is in bad trouble. It is easy for me to say why the preempted alternative is not a cause: the effect does not depend on it. My problem is to say why the preempting cause *is* a cause, when the effect does not depend on it either. (A regularity analysis of causation has the opposite problem: why is the preempted alternative not a cause, when it is part of a set of conditions jointly sufficient for the effect?)

I subdivide preemption into *early* and *late*. In early preemption, the process running from the preempted alternative is cut off well before the main process running from the preempting cause has gone to completion. Then somewhere along that main process, not too early and not too late, we can find an intermediate event to complete a causal chain in two steps from the preempting cause to the final effect. The effect depends on the intermediate, which depends in turn on the preempting cause. (Or, in cases with more than one preempted alternative, we might need more steps.) We have a causal chain of stepwise dependence between the cause and the effect, even if not dependence *simpliciter*; and since causation is transitive, we take the ancestral of dependence. Thus I say that *c* is a cause of *e* if there is a sequence *c*, . . . , *e* of events, consisting of *c* and *e* and zero or more intermediates, with each event in the sequence except the first depending on the one before. (Normally all these events would be distinct, and in temporal order; but I do not require this. See Postscript F, below.)

This is the variety of preemption that I discussed in the paper. To illustrate it, let us have another system of neurons.
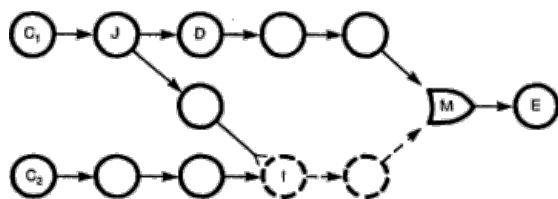
Besides stimulatory synapses from one neuron to another, as before, we now have an inhibitory synapse as well (shown by a backward arrowhead). A neuron normally fires if stimulated, but not if it is inhibited

at the same time. Neurons C_1 and C_2 fire; thereby starting two processes of firing which make their separate ways toward neuron E. The main process, which begins with the firing of C_1, goes to completion. But the alternative process, which begins with the firing of C_2, is cut short: because neuron I is inhibited, the neurons shown dotted never fire. There is also a branch process, diverging from the main process. The junction event where it diverges is the firing of neuron J. It is this branch process that cuts off the alternative process by inhibiting neuron I. The main and alternative processes—the one actual, the other partly unactualized—merge with the firing of neuron M; and proceed thence to the final effect, the firing of neuron E.

Thus the firing of C_2 is the preempted alternative. It is not a cause of the firing of E because there is no direct dependence, and neither is there any stepwise dependence via an intermediate. The firing of C_1 is the preempting cause. The firing of D is our intermediate event. It depends counterfactually on the firing of C_1; the firing of E depends on it; and thereby we have our two-step chain of dependence from the preempting cause to the effect. For by the time of the firing of D, the alternative process was already doomed. The alternative process was doomed as soon as neuron J fired; though it was not yet cut off, the branch process that was going to cut it off had already diverged from the main process. So if the firing of D had not occurred, *both* processes would have failed, and the firing of E also would not have occurred.

Don't say that if D had not fired, that would mean that it had not been stimulated, and that would mean that the neurons to its left on the main process would not have fired, and so neuron I would not have been inhibited, and so the alternative process would have gone to completion and E would have fired after all. That is backtracking; and backtracking counterfactuals, however legitimate in other contexts, are out of place in tracing causal dependence. (See "Counterfactual Dependence and Time's Arrow" in this volume.) Of course it is not just to deal with early preemption that we must avoid backtracking; as is explained in the paper, the avoidance of backtracking is needed also to solve the problems of effects and of epiphenomena.

We have some choice which event goes along the main process to take as our intermediate. The firing of J comes too early: the effect does not depend on it, since without it the alternative process would not have been cut off. The firing of M comes too late: it lies on the unactualized alternative process as well as on the main process, and so does not depend on the preempting cause. But anything in between would do. What makes the solution possible is that there exists some

intermediate event in the gap between too early and too late. And so it is, generally, in cases of early preemption. Thus we distinguish the genuine cause from its preempted alternative, as we should, even though either one by itself would have sufficed to cause the effect. [24]

[24] See Bruce LeCatt, "Censored Vision," *Australasian Journal of Philosophy* 60 (1982): 158–62, for further examples of stepwise dependence in cases of early preemption.

Late preemption is harder. Our solution cannot succeed unless there is a sufficient gap between too early and too late; if not-too-early is already too late, there is no place for an intermediate event to complete a chain of stepwise dependence.

There are two far-fetched ways in which this problem might arise. The first way involves action at a temporal distance. Suppose that in our previous example, we remove all the neurons between J (too early) and M (too late).
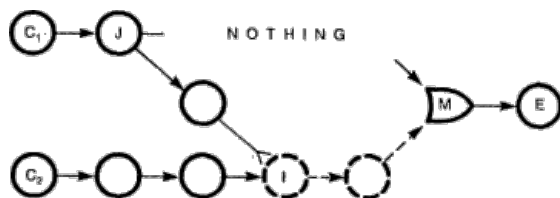
In their place, suppose we have some law of delayed action that directly connects the firings of J and M. Iff J fires, then M fires a certain time later (as in the original example) but that is absolutely all there is to it—there is no connection between the two neurons, and no continuous causal process between their two firings. That is possible, I take it, though it goes against what we take to be the ways of this world. In such a case, we have no intermediate event to complete our chain of dependence.

The second way involves infinite multiple preemption. We have infinitely many preempted alternatives, and infinitely many cut-off alternative processes. Suppose for simplicity that the main process and its unactualized alternatives merge only at the final effect. (Otherwise the problem would be the same, but with the point of merging in place of the final effect.) Then any other event on the main process is not too late to depend on earlier events along that process. The problem is to

find an intermediate event that is not too early to take the penultimate place in our chain of stepwise dependence—that is, to find an event on which the final effect depends. Such an event has to come late enough that by the time it occurs, all of the infinitely many alternative processes are doomed. Any one of the alternative processes is eventually doomed, so there is an event that comes late enough so far as it is concerned. Likewise for any finite set. But since there are infinitely many alternatives, there may be no event before the final effect that comes after *all* the alternative processes are doomed. Suppose one of them is doomed 128 seconds before the final effect, another only 64 seconds before, another only 32 seconds before,. . . . Then at no time before the final effect are all of them doomed. Then there is no intermediate event on which the final effect depends. Our causal chains of stepwise dependence can get as close as we like to the final effect, but they never can reach it. Then there is no stepwise dependence between the effect and what seems to be its preempting cause. [25]

[25] See William K. Goosens, "Causal Chains and Counterfactuals," *Journal of Philosophy* 76 (1979): 489–95 🖳 Link▸ .

I do not worry about either of these far-fetched cases. They both go against what we take to be the ways of this world; they violate the presuppositions of our habits of thought; it would be no surprise if our common-sense judgements about them therefore went astray —spoils to the victor! Common sense does judge them to be cases of causal preemption, in which what seems to be a preempting cause is indeed a cause, despite the lack of either direct or stepwise dependence. But an analysis that disagrees may nevertheless be accepted. It would be better to agree with common sense about these cases, to be sure, but that is not an urgent goal.

Unfortunately there is another variety of late preemption, quite commonplace and not at all far-fetched; and there it *is* an urgent goal to agree with common sense. Again we have what seems to be a preempting cause, hence a cause *simpliciter*, but no dependence and no stepwise dependence. Here my analysis seems to be in trouble. These are cases in which an alternative process is doomed only when the final effect itself occurs. The alternative is cut off not by a branch process that diverges from the main process at a junction event before the effect is reached, but rather by a continuation of the main process beyond the effect. Shooting a man stalked by seven other gunmen would be a case of this kind, if it is a case of redundant causation at all, and if the other gunmen desist only when they see him dead. Another case would be

this system of neurons. Again we start with the simultaneous firings of neurons C_1 and C_2, which redundantly cause the firing of E.
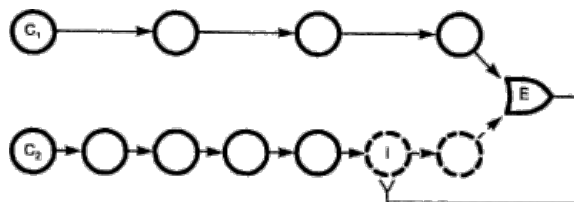
I ignored such cases when I wrote the paper, and for many years afterward. My reason must have been that there is a ready-made solution: fragility of the effect. If the alternative process is only doomed by the effect itself, and if at the time of the effect it is not yet complete, then the alternative process must run more slowly. So if it had been left to produce the effect, the effect would have been delayed. Without the firing of neuron C_1 (the seeming main cause) the firing of neuron E would have been delayed by the time it takes for three extra neurons to fire; if you had not shot the man on Tuesday morning, he would not have died until Wednesday evening; and so on, for all such cases. We can devise cases in which the delay is very short, but we can never get rid of it altogether. (Or not without resort to instantaneous or backward causation. But then the case becomes far-fetched, not worrisome, spoils to the victor.) If the effect is taken to be fragile, then the delay would suffice to give us a numerically different event instead of the effect that actually occurred. We would have causal dependence without redundancy, thus agreeing with common sense that your shooting the man on Tuesday, or the firing of C_1, or whatever, is indeed a cause.

But my reason for ignoring these cases was a bad reason, because the ready-made solution is a bad solution. Fragility of the effect is no better as a remedy for these cases of late preemption than it is as a remedy for redundant causation generally. To deal with all the cases, including those where the delay is very short and there is not much difference in manner to go with it, we need extreme standards of fragility; uniformly extreme standards are no good because they will give us lots of spurious causal dependence; so we need a double standard; and that might be workable, for all we know, but we don't know how to make it work. There are two problems. One is that a double standard must be principled. We need some definite rule to tell us when we should raise the standard: when is dependence among fragile versions relevant,

and when is it not, to causation among the original robust events? The second problem is that a stringent standard may give the wrong answer. Let c_1 be a preempting cause of e, and let c_2 be the preempted alternative, in a case of late preemption. Without c_1, e would have been delayed; and so a more fragile version of e would not have occurred at all. So far, so good. But it may also be that some side effect of c_2 substantially influences the time and manner of e; in which case, unfortunately, a version of e that is fragile enough to depend on c_1 may depend on c_2 as well. Indeed, it may take more fragility to give us the dependence on c_1 that yields the right answer than it does to give us the dependence on c_2 that yields the wrong answer. Though I don't reject the fragility approach out of hand, I don't see how to make it work. [26]

[26] I am indebted to discussion with D. H. Rice, who has persuaded me that it would be premature to give up on fragility solutions without a good deal of further investigation.

So I am inclined to prefer a different solution, though it is more of a departure from my original analysis in the paper.

Leaving the problem of late preemption in abeyance, consider this question. Suppose we have processes—courses of events, which may or may not be causally connected—going on in two distinct spatiotemporal regions, regions of the same or of different possible worlds. Disregarding the surroundings of the two regions, and disregarding any irrelevant events that may be occurring in either region without being part of the process in question, what goes on in the two regions is exactly alike. Suppose further that the laws of nature that govern the two regions are exactly the same. Then can it be that we have a causal process in one of the regions but not the other? It seems not. Intuitively, whether the process going on in a region is causal depends only on the intrinsic character of the process itself, and on the relevant laws. The surroundings, and even other events in the region, are irrelevant. Maybe the laws of nature are relevant without being intrinsic to the region (if some sort of regularity theory of lawhood is true) but nothing else is.

Intuitions of what is intrinsic are to be mistrusted, I think. They too often get in the way of otherwise satisfactory philosophical theories. Nevertheless, there is some slight presumption in favor of respecting them. Let us see where this one leads us.

A process in a region may exhibit a pattern of counterfactual dependence that makes it causal, according to my original analysis. Its later parts may depend counterfactually on its earlier parts (later and earlier in time, normally, but all I require is that there be dependence with

respect to some order); and in particular, its last event may depend on its first. (We will provide for stepwise dependence later.)

Now suppose that some process in some region does not itself exhibit this pattern of dependence; but suppose that in its intrinsic character it is just like processes in other regions (of the same world, or other worlds with the same laws) situated in various surroundings. And suppose that among these processes in other regions, the great majority—as measured by variety of the surroundings—do exhibit the proper pattern of dependence. This means that the intrinsic character of the given process is right, and the laws are right, for the proper pattern of dependence—if only the surroundings were different, and different in any of many ways. According to my original analysis, this process is nevertheless not causal. Thanks to its special bad surroundings, it is a mere imitation of genuine causal processes elsewhere. But that goes against our motivating intuition.

So we might extend the analysis. Suppose that there exists some actually occurring process of the kind just described, and that two distinct events c and e are the first and last in that process. Then let us say that e *quasi-depends* on c. We might wish to count that as one kind of causation, based derivatively on counterfactual dependence even though there is no dependence between those two events themselves. As before, we must take an ancestral to ensure that causation will come out transitive; thereby providing not only for chains of stepwise dependence, but also for chains of stepwise quasi-dependence, or mixed chains. To this end we could redefine a *causal chain* as a sequence of two or more events, with either dependence or quasi-dependence at each step. And as always, one event is a *cause* of another iff there is a causal chain from one to the other.

That would solve the problem of late preemption, both in the commonplace cases that worry me and in the far-fetched cases that do not. For the problem is that we seem to have a causal process starting with a preempting cause, and ending with the final effect; and yet this process does not exhibit the proper pattern of counterfactual dependence, not even if we count stepwise dependence. Segments of it do exhibit dependence, but we cannot patch these segments together to make a chain that reaches all the way to the effect. What spoils the dependence is something extraneous: the presence alongside the main process of one or more preempted alternatives. Without them, all would be well. Hold fixed the laws but change the surroundings, in any of many ways, and we would have the dependence that my original analysis requires for causation. But as is, we have quasi-dependence instead of

dependence. So if we extend the analysis, and allow causation by quasi-dependence, that solves our problem. We then can agree with common sense that we have genuine preemption, and genuine causation by the preempting cause. [27]

[27] If we admit causation by quasi-dependence, it would be nice if that could buy us some simplification elsewhere. Could we perhaps drop the part of the analysis in which we take an ancestral to ensure that causation turns out transitive? I think not, in view of a case suggested by John Etchemendy. Suppose we have a case of preemption with this peculiarity: there is no way, given the laws of nature, that the preempting cause could fail to have been accompanied by the preempted alternative. Any lawful way of producing one must produce the other as well. It seems that we have a main causal process running from a preempting cause c to its final effect e. Because of the preemption, e does not depend directly on c. And neither do we have direct quasi-dependence: any process just like it, and under the same laws, must likewise have its dependence destroyed by preemption. The problem comes not from an accident of circumstances, but from the laws themselves. So if we admit causation by quasi-dependence but we do not take an ancestral, we still get the wrong answer. But take some intermediate event d along the main process from c to e, before the point where it merges with the alternative process. For the first step, we have causation by dependence: d does depend on c. For the second step, we may have causal dependence of e on d if the preemption is early. We may not, if the preemption is late; but even so, assuming that d could have been produced without also producing the preempted alternative, we at least have quasi-dependence of e on d. So we have a chain from c to d to e, with dependence or quasi-dependence at both steps. Then if we take an ancestral to ensure that causation comes out transitive, we get the right answer.

What if there is *no* intermediate that could lawfully have been produced without also producing a preempted alternative? That makes the case very peculiar indeed. It is central to the way we ordinarily think about preemption that we can regard the main and the alternative processes as distinct and separable. So if the laws forbid us to have even a part of the one process without the corresponding part of the other, that goes badly against our habitual presuppositions. If so, such common sense opinions as we may have need not be respected—spoils to the victor.

The extended analysis, which allows causation by quasi-dependence, is more complicated than my original analysis; and it is less purely a counterfactual analysis, though of course counterfactual dependence still plays a central role. The complication would be objectionable if it were just a hoky gimmick to deal with late preemption, but it is not just that. For what it is worth, we also have independent motivation in the intuition of intrinsicness. While I would still welcome a different solution to the problem of late preemption, within my original analysis, I now think that the extended analysis may well be preferable.

This completes my discussion of preemption. I now turn to the other variety of redundant causation: overdetermination, with nothing to break the symmetry between the redundant causes. When I wrote

the paper, I thought that all such cases were alike; that a counterfactual analysis would inevitably deny that the redundant causes in overdetermination are causes *simpliciter*; and that it did not matter much what the analysis said, since all such cases were spoils to the victor for lack of firm common-sense judgements.

All that is wrong. An important paper by Martin Bunzl changes the picture greatly. [28]

[28] "Causal Overdetermination," *Journal of Philosophy* 76 (1979): 134–50 🖥 Link ▶ .

Bunzl observes that when we examine stock examples of overdetermination in detail, we can very often find an intermediate event—call it a *Bunzl event*—that satisfies two conditions. First, the Bunzl event is jointly caused, without redundancy, by the same events that are redundant causes of the final effect. Second, the Bunzl event seems clearly to be a cause (often a preempting cause) of the final effect. Cases of overdetermination are not all alike, because there are different kinds of Bunzl events (at least three) and also because there are some possible cases, far-fetched perhaps, with no Bunzl events at all. A counterfactual analysis does not deny that the redundant causes are causes *simpliciter* of the final effect, provided it can agree that they are causes of a Bunzl event and that the Bunzl event in turn is a cause of the effect. The cases should not all be left as spoils to the victor, because once a Bunzl event is noticed, it becomes clear to common sense that we have genuine causation.

One kind of Bunzl event is a preempting cause in a case of late preemption. This system of neurons illustrates it. Here B is an especially lethargic neuron. It will not fire if singly stimulated, but it will if doubly stimulated.
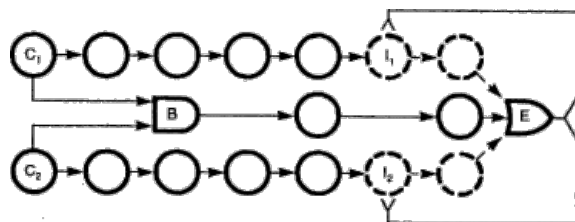
As usual, the simultaneous firings of C_1 and C_2 are redundant causes of the firing of E. But also they are joint causes, without

redundancy, of a Bunzl event: namely, the firing of B. And that is a preempting cause of the final effect. The preemption is late: the two alternative processes,

those that run from the firings of C_1 and C_2 taken separately, are cut off only because of the effect itself. It is the firing of E itself that inhibits neurons I_1 and I_2. We must apply whatever solution we favor for late preemption generally. If somehow we had a double standard of fragility, we might say that the firing of E is extremely fragile, and would not have occurred (though E would still have fired) without the firing of B. Or, probably better, we might say that despite a lack of direct or stepwise dependence, we have causation by quasi-dependence. Thanks to the intrinsic character of the course of events running from the firing of B to the firing of E, we would have had counterfactual dependence if just such a course of events had occurred under the same laws but in any of various different surroundings. According to the extended analysis, that means that the firing of B is a cause of the firing of E. Either way, we say as we should that the firing of B causes the firing of E; and therefore, by transitivity, the firing of C_1 and C_2 are both causes of E.

Not only is the firing of B a preempting cause; we can also think of the entire course of events as a case of *self*-preemption. The firings of C_1 and C_2, our redundant causes, both preempt and are preempted. Taken together as joint causes, they preempt themselves taken separately.

A second kind of Bunzl event is a preempting cause in a case of early preemption, as in the following system of neurons.
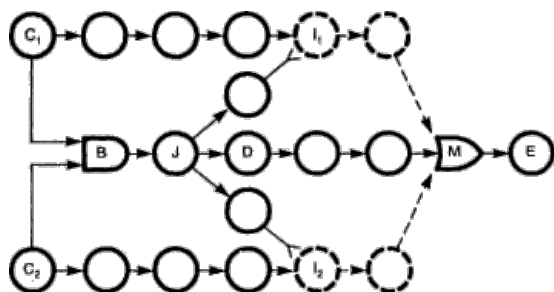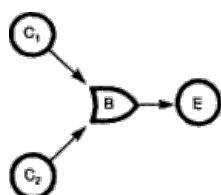

Again the simultaneous firings of C_1 and C_2 are redundant causes of the firing of E, and joint causes without redundancy of the firing of the lethargic neuron B. As in simpler cases of early preemption, the firing of D completes a chain of stepwise dependence: it depends on the firing of B, and in turn the firing of E depends on the firing of D. So there is a three-step causal chain of dependence from the firing of C_1 to the firing

of E; and likewise from the firing of C_2. More simply, there is a two-step chain, since the firing of D also depends directly on the firing of C_1, and likewise on the firing of C_2. The firing of B is a Bunzl event; so is the firing of D; and so are various other intermediate events on the chain. Again we have self-preemption by our redundant causes: the firings of C_1 and C_2 taken jointly preempt themselves, taken separately.

This looks complicated. But just the same sort of early self-preemption can happen in much simpler cases of overdetermination, as follows.

The third kind of Bunzl event is a fragile intermediate. Earlier, we considered a case of fragility of the effect, involving a neuron that would fire vigorously if doubly stimulated, feebly if singly stimulated. We considered that under moderate and reasonable standards of fragility, hence without any problematic double standard, we might say that the vigorous firing and the feeble firing would differ enough in manner to make them numerically different events. If we place the fragile vigorous firing as an effect, what we have is not redundant causation at all. But if we place it as an intermediate, it can be the Bunzl event in a case of overdetermination. Here is such a case, with B as the neuron that may fire either vigorously or feebly.

The vigorous firing of B that actually occurs depends on both of the simultaneous firings of C_1 and C_2. Without either one of these causes

it would not have occurred. The feeble firing of B that would have occurred with only one of them would not have been the same event. But also the firing of E depends on the firing of B. So each of our redundant causes is connected to the final effect by a two-step causal chain of dependence. Not by direct dependence: if only one of C_1 and C_2 had fired, so that B fired feebly, E would still have been stimulated and its firing would have been very little different. This is not a case that can be treated by fragility of the effect, or not under moderate standards of fragility.

(My solution depends on assuming that if the intermediate event—the vigorous firing of B—had not occurred, then B would not have fired at all. It isn't that the vigorous firing would have been replaced by

a feeble firing, differing only just enough not to be numerically the same. That may seem to go against a similarity theory of counterfactuals—wouldn't the minimal change to get rid of an event be one that replaces it with a barely different event? Not so; a similarity theory needn't suppose that just any sort of similarity we can think of has nonzero weight. It is fair to discover the appropriate standards of similarity from the counterfactuals they make true, rather than *vice versa*. (See "Counterfactual Dependence and Time's Arrow" in this volume.) And we certainly do not want counterfactuals saying that if a certain event had not occurred, a barely different event would have taken its place. They sound false; and they would make trouble for a counterfactual analysis of causation not just here, but quite generally.)

The case looks simpler than the self-preemption cases above; but it is really much the same. The process from the redundant causes jointly through the vigorous firing to the effect goes to completion. The two alternative processes from the redundant causes taken singly through the feeble firing to the effect are cut short when the feeble firing does not occur. The feeble firing is prevented by the double stimulation of **B**, and that is an event in the main process.

Still there is one important difference from previous cases. When we have a fragile intermediate, as opposed to the sorts of Bunzl events considered above, there is room for serious indeterminacy. Just as our vague and shifty standards of fragility may leave it unsettled whether we have a fragile effect, so they may leave it unsettled whether we have a fragile intermediate. Then they may leave it unsettled whether we have overdetermination with or without a Bunzl event. If that is what decides whether the redundant causes are causes *simpliciter*, that question too may have no right answer.

So I turn to the last variety of redundant causation: overdetermination without a Bunzl event, including doubtful cases when taken under standards of fragility that give no relevant fragility either in the effect or in the intermediates. According to my original analysis, the redundant causes in such a case are not causes *simpliciter*, because there is neither direct nor stepwise dependence. But the extended analysis would disagree. There is quasi-dependence of the effect on each of the two redundant causes, and if we allowed causation by quasi-dependence, that would make the redundant causes count as genuine causes of the effect.

Also, the original analysis will say that in cases where it is doubtful whether there is a fragile effect or intermediate, then it is likewise doubtful whether the redundant causes are causes *simpliciter*. Whereas

the extended analysis would say that in such cases the redundant causes are causes, though the reason why is left doubtful. The first analysis would be better suited to explain indecision and controversy, the second would be better suited to explain positive judgements.

I used to think that all cases of overdetermination, as opposed to preemption, could be left as spoils to the victor; and that is what I still think about these residual cases. All the more so, given Bunzl's discussion of what we find when we look at realistic cases in microscopic detail, without simplifying idealizations. For it seems that cases without Bunzl events require phenomena with perfectly sharp thresholds, whereas thresholds under the laws of this world are imperfectly sharp. Thus I am content to say that these cases may go one way or the other. The decision will depend on what strategy emerges as victor in the cases that really matter—namely, the commonplace cases of late preemption.

I should dispel one worry: that if we ever decline to count redundant causes as genuine causes, then we will be left with gaps in our causal histories—no cause at all, at the time when the redundant causes occur, for a redundantly caused event. That is not a problem. For consider the larger event composed of the two redundant causes. (I mean their mereological sum. *Not* their disjunction—I do not know how a genuine event could be the disjunction of two events both of which actually occur. It would have to occur in any region where either disjunct occurs. Hence it would have to occur twice over in one world, which a particular event cannot do. See "Events" in this volume.) Whether or not the redundant causes themselves are genuine causes, this larger event will be there to cause the effect. For without it—if it were completely absent, with neither of its parts still present, and not replaced by some barely different event—the effect would not occur. For *ex hypothesi* the effect would not occur if both redundant causes were absent, and to suppose away both of them is

just the same as to suppose away the larger event that is composed of them.

F. Self-Causation

My requirement that cause and effect be distinct applies to causal dependence, but not to causation generally. Two events are distinct if they have nothing in common: they are not identical, neither is a proper part of the other, nor do they have any common part. Despite the truth of the appropriate counterfactuals, no event depends causally

on itself; or on any other event from which it is not distinct. However, I do allow that an event may cause itself by way of a two-step chain of causal dependence: c depends on d which depends in turn on c, where d and c are distinct. Likewise for longer closed causal loops; or for loops that lead from an event back not to itself but to another event from which it is not distinct. Thus I have taken care not to rule out the sort of self-causation which appears in time-travel stories that I take to be possible. (See "The Paradoxes of Time Travel" in this volume.)

But no event can be self-caused unless it is caused by some event distinct from it. Indeed, no event can be caused at all unless it is caused by some event distinct from it. Likewise no event can cause anything unless it causes some event distinct from it.

Suppose we think of the entire history of the world as one big event. It is not caused by any event distinct from it; else that distinct event both would and would not be part of the entire history. Likewise it does not cause any event distinct from it. So it has no causes or effects at all. Not as a whole, anyway. Its parts, of course, do all the causing there is in the world.

Some philosophers wish to believe only in entities that have some causal efficacy. [29]

[29] For instance, see D. M. Armstrong, *Universals and Scientific Realism*, Volume I (Cambridge: Cambridge University Press, 1978), pp. 128–32.

Either they must reject such totalities as the big event which is the whole of history, or else they should correct their principle. They might admit those inefficacious things that could have been efficacious if, for instance, there had been more of history than there actually was. Or, more simply, they might admit those inefficacious things that are composed entirely of efficacious parts.