# Causation and Counterfactuals

edited by John Collins, Ned Hall, and L. A. Paul

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Times New Roman on 3B2 by Asco Typesetters, Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Causation and counterfactuals / edited by John Collins, Ned Hall, and L. A. Paul.
p. cm. — (Representation and mind)
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-03317-8 (hc. : alk. paper) — ISBN 0-262-53256-5 (pb. : alk. paper)
1. Causation. 2. Counterfactuals (Logic). I. Collins, John David. II. Hall, Edward J. (Edward Jonathan), 1966– . III. Paul, L. A. (Laurie Ann), 1966– . IV. Series.
BD541.C192 2004
122—dc22 2003060612

10 9 8 7 6 5 4 3 2 1

To the memory of David Lewis

# 3 Causation as Influence

David Lewis

## 1 Lecture I

My second paper in my first philosophy course defended a counterfactual analysis of causation. I've been at it, off and on, ever since. But it's obvious that the simplest counterfactual analysis breaks down in cases of redundant causation, wherefore we need extra bells and whistles. I've changed my mind once more about how those bells and whistles ought to work.

This paper mostly presents the latest lessons I've learned from my students. Under the customs of the natural sciences, it should have been a joint paper, the coauthors being (in alphabetical order) John Collins, Ned Hall, myself, L. A. Paul, and Jonathan Schaffer. But under the customs of philosophy, a paper is expected to be not only a report of discoveries, but also a manifesto; and, happily, the five of us have by no means agreed upon a common party line. So, while I'm much more than usually indebted to the work of Collins, Hall, Paul, and Schaffer, they cannot be held responsible for the position I have reached.

### 1.1 Why Seek a Counterfactual Analysis?

The best reason to persist in trying to make a counterfactual analysis of causation work is that the difficulties that confront rival approaches seem even more daunting.

It is not a foregone conclusion that causation requires analysis at all. Is there, perhaps, an unanalyzable relation of singular causation, which we know by perceptual acquaintance, and which we are therefore in a position to refer to and think about? It might indeed turn out that this relation can be identified with some relation already familiar to us either from physics or from metaphysical speculation; but if so, that identification would be a physical or metaphysical hypothesis, not a matter for a priori conceptual analysis.[1]

Hume, of course, taught that we never perceive causation, but only repeated succession. But it's famously difficult to draw the line between what's true according to perceptual experience all by itself and what's true according to a system of beliefs shaped partly by perceptual experience and partly by previous beliefs. The boot comes forward and touches the ball, and straightway the ball flies off through the goalposts. Do I see that the one thing causes the other? Or do I infer it from what I do see, together with my background knowledge about the ways of the world? I don't

Originally published in a special issue of the *Journal of Philosophy* (April 2000). Minor revisions have been made for consistency.

know, and I don't know how to find out. So I'm in no position to deny that in such a case I'm perceptually acquainted with an instance of a causal relation; and thereby acquainted as well with the relation that is instantiated.[2]

I'm acquainted with *a* causal relation—not with *the* causal relation. Causal relations are many and various, and no amount of watching the footy will acquaint me with all the causal relations there are, let alone all the causal relations there might have been. And yet I seem to have picked up a *general* concept of causation, applicable to all different kinds of causation, and applicable even to kinds of causation never found in our own world. That's the real problem, even if I concede *pace* Hume that I sometimes perceive causation.

If ever I perceive causation, I perceive it when I watch the footy; or, to take the customary example, if I watch the motions of billiard balls. But the causal mechanism whereby a dinner too low in carbohydrate causes low blood sugar is utterly different. The causal mechanism whereby our former congressman helped cause his own defeat by literally singing the praises of Kenneth Starr is different again. And so on, and so forth; not quite ad infinitum if we limit ourselves to actuality. But we should not limit ourselves to actuality, given that we can perfectly well understand fantasies, or theologies, in which causation works by magical mechanisms entirely alien to the world of our acquaintance. We are not perceptually acquainted with each and every one of all these different actual and possible causal relations. If there is a single causal relation, either it is a far from natural relation, a gruesomely disjunctive miscellany, and so not the sort of relation we can become acquainted with by being acquainted with some few of its disjuncts; or else the many disjuncts have something in common. I think conceptual analysis is required to reveal what it is that all the actual and possible varieties of causation have in common.

A parallel objection applies to the "Canberra plan" for causation. The plan is that we first elicit, Meno-fashion, the folk theory of causation that we all implicitly hold. That done, we can define causation as whatever it is that comes closest, and close enough, to occupying the role specified by our folk-theoretical platitudes. We require conceptual analysis of the job description, so to speak; but not of the actual occupant of the role specified thereby.[3,4] We leave it open that the role may be occupied by different relations in different possible worlds, thereby explaining how our concept of causation applies to causation by possible mechanisms alien to actuality. But the problem of the many diverse actual causal mechanisms, or more generally of many diverse mechanisms coexisting in any one world, is still with us. If causation is, or might be, wildly disjunctive, we need to know what unifies the disjunction. For one thing the folk platitudes tell us is that causation is one thing, common to the many causal mechanisms.

The problem becomes especially acute when we remember to cover not only causation of positive events by positive events, but also causation by absences, causation of absences, and causation via absences as intermediate steps. The most fundamental problem is that absences are unsuitable relata for any sort of causal relation, by reason of their nonexistence. This is everyone's problem. It is not to be dodged by saying that causation involving absences is really "causation*," a different thing from genuine causation—call it what you will, it still needs to be part of the story. It is my problem too, and I shall return to it; but in the meantime, let the missing relata objection join the miscellany objection as reasons to think that acquaintance with "the" causal relation, or characterization of "it" as the occupant of a role, are not workable rivals to a conceptual analysis of causation.

If we are convinced of that, one rival to the counterfactual analysis remains standing. That is the analysis that says, roughly, that a cause is a member of a set of conditions jointly sufficient, given the laws of nature, for the effect (or perhaps for a certain objective probability thereof). (See White 1965, pp. 56–104; and Mackie 1965.) This deductive-nomological analysis is descended from Hume's constant conjunction theory, just as our counterfactual analysis is descended from Hume's offhand remark that "if the first object had not been, the second never had existed."[5]

However, we don't want to count *C* as a cause of *E* just because *C* belongs to some sufficient set or other for *E*: A sufficient set remains sufficient if we add irrelevant rubbish, and *C* might be exactly that. *C* should belong to a *minimal* sufficient set, and that is not easy to define. It won't work just to say that no condition in the set may be deleted without rendering the remainder insufficient: that can be circumvented by mingling relevant and irrelevant information in such a way that each item in the set contains some of each. I might suggest appealing to a counterfactual: we want our sufficient set to consist of items without which the effect would not have occurred. (That doesn't work as it stands, but it's at least a step in the right direction.) But now we have departed from the deductive-nomological analysis in the direction of a counterfactual analysis.

Another difficulty is that it can perfectly well happen that an effect is a member of a minimal jointly sufficient set for its cause; or that one effect of a common cause is a member of a minimal jointly sufficient set for another. The falling barometer supposedly causes the low pressure; or the falling barometer supposedly causes the storm. Even if we were willing to declare a priori that no cause ever precedes its effect, that would be no solution. The falling barometer *does* precede the storm. I know of no solution to these familiar difficulties within the confines of a purely deductive-nomological analysis of causation.

The final rivals to a satisfactory counterfactual analysis of causation are the *un*satisfactory counterfactual analyses. Long, long ago, I thought it would suffice to say that event *C* is a cause of event *E* iff *E* depends counterfactually on *C*; iff, if *C* had not occurred, *E* would not have occurred. But this turns out to need qualifications before we have even a sufficient condition for causation.

First. We need the right kind of relata. *C* and *E* must be distinct events—and distinct not only in the sense of nonidentity but also in the sense of nonoverlap and nonimplication. It won't do to say that my speaking this sentence causes my speaking this sentence; or that my speaking the whole of it causes my speaking the first half of it, or vice versa; or that my speaking it causes my speaking it loudly, or vice versa. Nor should *C* and *E* be specified in an overly extrinsic way; it won't do to say that events a third of a century ago caused me to speak this sentence in the place where once I was a student. (Though those events did cause my speaking simpliciter.) See Kim (1973b) and Lewis (1986d).

Second. We need the right kind of counterfactual conditionals. Why can't we say, given the laws connecting barometer readings and air pressure, that if the barometer hadn't fallen, that would have been because the pressure wasn't low? Why can't we then conclude that if the barometer hadn't fallen, there wouldn't have been a storm? Yet if we say such things, why doesn't our counterfactual analysis fail in just the same way that the deductive-nomological analysis did?—I agree that we're within our linguistic rights to assert these backtracking, or back-and-then-forward, counterfactuals. But they are out of place in the context of establishing causal connections. Here the much-bemoaned flexibility of counterfactual conditionals is our friend. When we imagine Caesar in command in Korea, we have a choice: We can hold fixed Caesar's military knowledge, or we can hold fixed the weaponry of the Korean war. Likewise when we imagine the barometer not falling, we have a choice: we can hold fixed the previous history, or we can hold fixed the lawful connections between that history and what the barometer does. For purposes of analyzing causation, our policy in all such cases must be to prefer the first choice to the second. If need be, we hold history fixed even at the price of a miracle (see my 1979a).

Now I think our oversimple counterfactual analysis succeeds in characterizing one kind of causation. But other kinds are omitted. We have a sufficient, but not a necessary, condition for causation.

For one thing, we usually think that causation is transitive: if *C* causes *D*, which in turn causes *E*, it follows that *C* causes *E*. That is why we can establish causal connections by tracing causal chains. But we have no guarantee that the relation of counterfactual dependence will be invariably transitive. (Shortly we shall see how its transitivity can fail.) So we need to provide for causation not only by direct dependence, but by chains of stepwise dependence. We can do so by defining causation as the ancestral of dependence (see my 1986b).

But that still does not suffice to capture all cases of causation. We have at least three items of unfinished business. Probabilistic causation, preemptive causation, and causation by, or of, absences are not yet fully covered. Here I shall mostly be discussing the second and third topics.

### 1.2 Probabilistic Causation

I have little to say about probabilistic causation. Not because I don't believe in it: More likely than not, our world is so thoroughly indeterministic that most or all of the causation that actually takes place is probabilistic. Whether our world is governed by indeterministic laws is settled neither by the Moorean fact that we make free choices nor by a priori principles of sufficient reason. Rather, it is a contingent question of theoretical physics. If the best explanation of quantum phenomena requires spontaneous collapses of the wave function, then we should believe in widespread indeterminism. If Bohmian mechanics is a better explanation, we should believe that our world is deterministic after all. Either way, there is plenty of causation in the world. Those who believe in widespread indeterminism still ascribe causal connections. It would be preposterous to deny that the connections they ascribe deserve the name they are given. Therefore chancy events can be caused. They can be caused even when their causes do not make them inevitable, and do not even make them highly probable. They can be caused even when they have some slight chance of occurring spontaneously.

The probabilistic counterpart of the simplest sort of counterfactual dependence is, roughly, probability-raising. $C$ occurs, $E$ has a certain chance (objective single-case probability) of occurring, and as it happens $E$ does occur; but without $C$, $E$'s chance would have been less. Likewise, the more complicated patterns of counterfactual dependence that I shall be discussing later also come in probabilistic versions.

I used to think (substantial) probability-raising could simply take the place of all-or-nothing counterfactual dependence in an analysis of causation (see my 1986b, pp. 175–180). But there is a problem: Not all probability-raising counts. One terrorist places an unreliable bomb—a genuinely indeterministic device—on Flight 13; another terrorist places an unreliable bomb on Flight 17. As it happens, the bomb on Flight 13 goes off and the bomb on Flight 17 doesn't. The *Age* runs a headline: "Airline bomb disaster." The headline would have been just the same if it had been the bomb on Flight 17 that went off, or if it had been both. So the bomb on Flight 17 raised the probability of the headline, but certainly didn't cause it. We want to say that the raising that counts is the raising of the probability of the causal chain of events and absences whereby the effect was actually caused. Raising the probability of some unactualized alternative causal chain leading to the same effect doesn't count. But it would be circular to say it that way within an analysis of causation. I hope there is some noncircular way to say much the same thing, but I have none to offer.[6]

That said, let us set aside the probabilistic case. The proper treatment of causation in a deterministic world will give us difficulties enough. Those same difficulties would reappear for causation under indeterminism, and I hope the same solutions would apply.

### 1.3 Preemption Revisited

It sometimes happens that two separate potential causes for a certain effect are both present; and either one by itself would have been followed by the effect (or at least by a raised probability thereof); and so the effect depends on neither. Call any such situation a case of *redundant causation*. (For short: *redundancy*.) Some cases of redundancy are symmetrical: Both candidates have an equal claim to be called causes of the effect. Nothing, either obvious or hidden, breaks the tie between them. It may be unclear whether we ought to say that each is a cause or whether we ought to say that neither is a cause (in which case we can still say that the combination of the two is a cause). But anyway it is out of the question to say that one is a cause and the other isn't. Because it's unclear what we want to say, these symmetrical cases are not good test cases for proposed analyses of causation. Set them aside.

Other cases are asymmetrical. It's very clear what we want to say: One of the two potential causes did cause the effect, the other one didn't. Call the one that did the causing a *preempting* cause of the effect. Call the other one a *preempted* alternative, or *backup*.

When our opinions are clear, it's incumbent on an analysis of causation to get them right. This turns out to be a severe test. The simplest sort of deductive-nomological analysis flunks: The preempted alternative is a member of a minimal jointly sufficient set for the effect, yet it is not a cause. The simplest sort of counterfactual analysis likewise flunks: The preempting cause is not a condition without which the effect would have been absent, yet it is a cause. Both these attempts fail because they treat the preempting cause and its preempted alternative alike, whereas we know very well that one is a cause and the other is not. A correct analysis will need to discern the source of the difference.

### 1.4 Trumping

I used to think that all cases of preemption were cases of *cutting*: cases in which, first, there is a completed causal chain (often, but not necessarily, spatiotemporally continuous) running from the preempting cause all the way to the effect; but, second, something cuts short the potential alternative causal chain that would, in the absence of the preempting cause, have run from the preempted alternative to the effect.[7] Some think so still, but I have learned better.[8]

The Sergeant and the Major are shouting orders at the soldiers. The soldiers know that in case of conflict, they must obey the superior officer. But, as it happens, there is no conflict. Sergeant and Major simultaneously shout "Advance!"; the soldiers hear them both; the soldiers advance. Their advancing is redundantly caused: If the Sergeant had shouted "Advance!" and the Major had been silent, or if the Major had shouted "Advance!" and the Sergeant had been silent, the soldiers would still have advanced. But the redundancy is asymmetrical: Since the soldiers obey the superior officer, they advance because the Major orders them to, not because the Sergeant does. The Major preempts the Sergeant in causing them to advance. The Major *trumps* the Sergeant.

We can speculate that this might be a case of cutting. Maybe when a soldier hears the Major giving orders, this places a block somewhere in his brain, so that the signal coming from the Sergeant gets stopped before it gets as far as it would have done if the Major had been silent and the Sergeant had been obeyed. Maybe so. Or maybe not. We don't know one way or the other. It is epistemically possible, and hence it is possible simpliciter, that this is a case of preemption without cutting.

If we forsake everyday examples, we become free to settle by stipulation that we have no cutting. We can stipulate, for instance, that the causal process in question works by action at a distance. Nothing goes missing when the process is preempted, because ex hypothesi there are no intermediate events to go missing. Here is such an example. Suppose the laws of magic state that what will happen at midnight must match the first spell cast on the previous day. The first spell of the day, as it happens, is Merlin's prince-to-frog spell in the morning. Morgana casts another prince-to-frog spell in the evening. At midnight the prince turns into a frog. Either spell would have done the job, had it been the only spell of the day; but Merlin's spell was first, so it was his spell that caused the transmogrification. Merlin's spell trumped Morgana's. Merlin's spell was a preempting cause, Morgana's was the preempted backup. But we stipulate also that the causal process from spell to transmogrification has no intermediate steps.[9]

## 1.5 Commonplace Preemption

Trumping shows that preemption does not require the cutting of a causal chain. Nevertheless, the most familiar variety of preemption does work by cutting. The causal chain from the preempting cause gets in first: it runs to completion, and the effect happens, while the chain from the preempted alternative is still on its way. The preempted chain is cut. The effect itself is what prevents its final steps.

Billy and Suzy throw rocks at a bottle. Suzy throws first, or maybe she throws harder. Her rock arrives first. The bottle shatters. When Billy's rock gets to where the bottle used to be, there is nothing there but flying shards of glass. Without Suzy's throw, the impact of Billy's rock on the intact bottle would have been one of the final steps in the causal chain from Billy's throw to the shattering of the bottle. But, thanks to Suzy's preempting throw, that impact never happens.

I used to call such cases as this "late preemption." (In hindsight, "late cutting" would have been a better name.) I meant to contrast them with "early preemption": easy cases in which we have, if not direct counterfactual dependence of the effect itself on the preempting cause, at least stepwise dependence. The effect depends on some intermediate event, which in turn depends upon the preempting cause. (Or we may have stepwise dependence through a longer chain of intermediates.) These are cases in which dependence is intransitive, but we get the right answer by defining causation as the ancestral of dependence.

There is a small industry devoted to solving the preemption problem under the presupposed premise that preemption always works by cutting (see, e.g., Ramachandran 1997a). However well such solutions may (or may not) work in the cases they were made for, they are not general solutions because they do not deal with trumping. We may have to rest content with a patchwork of solutions, different ones for different cases, but let us hope for something more ambitious.

## 1.6 Quasi-dependence Rejected

I used to think that late preemption (and maybe early preemption as well) could be handled by appealing to the intuitive idea that causation is an intrinsic relation between events[10] (except insofar as being subject to such-and-such laws of nature is an extrinsic matter, as I believe it to be). Take another case, actual or possible, which is intrinsically just like the case of Suzy throwing her rock at the bottle (and which occurs under the same laws), but in which Billy and his rock are entirely absent. In this comparison case, we have a causal chain from Suzy's throw to the shattering which *does* exhibit counterfactual dependence and which is an intrinsic duplicate of the actual chain from Suzy's throw with Billy present. (Near enough. Doubtless the presence of Billy and his rock makes some tiny difference to the gravitational force on Suzy's rock, and therefore some negligible difference to that rock's trajectory.) I thought: If being a causal chain is an intrinsic matter, then both or neither of the two chains that are intrinsic duplicates (and occur under the same laws) must be causal; but the comparison chain, which exhibits dependence, surely is a causal chain; so the actual chain, even though thanks to Billy it does not exhibit dependence, must be a causal chain too. I said that the actual chain exhibited *quasi-dependence*: it qualified as causal by courtesy, in virtue of its intrinsic resemblance to the causal chain in the comparison case.

Quasi-dependence was a bad idea, for five reasons.

First. Imagine that Suzy's and Billy's rock-throwing takes place in a world with laws just a little different from what we take to be the laws of our actual world: laws under which flying objects sometimes make little random jumps. Imagine also that Suzy's rock is an intrinsic duplicate of Billy's. Now consider the chain of events consisting of Billy's throw, the flight of Billy's rock up to but not including the time when it reaches the place where the bottle used to be, plus the impact of Suzy's duplicate rock on the bottle and the shattering of the bottle. Compare this chain with another chain of events in which Suzy is absent, Billy throws, his rock takes a little jump just before impact, it hits the bottle, and the bottle shatters. The original chain and the comparison chain are intrinsic duplicates (or near enough) under the same laws. The comparison chain exhibits counterfactual dependence. But now we're forced to conclude that the shattering quasi-depends on *Billy's* throw! (As well as on Suzy's.) And that's the wrong answer: just as in the original case, Billy's throw is not a cause of the shattering, but rather a preempted alternative.[11]

Second. The intrinsic character of causation is, at best, a parochial feature of our own possible world. It does not apply, for instance, to an occasionalist world in which God is a third party to all causal relationships whatever between natural events. And yet occasionalism certainly seems to be a genuine possibility. So if we aim at conceptual analysis, not just a contingent characterization of the causal connections that are found in this world of ours, we cannot assume a priori that causation is an intrinsic matter.

Third. Quasi-dependence gives the wrong answer in cases of trumping preemption. The trumped causal chain runs to completion; therefore it is an intrinsic duplicate (near enough) of an *un*trumped causal chain in a comparison case (under the same laws) which exhibits counterfactual dependence. This reinforces our previous conclusion that quasi-dependence fails in some other possible worlds, for instance the world in which Merlin's spell trumps Morgana's. But worse, it may mean that the intrinsic character of causation is an overhasty generalization even about the causation that happens in our own world. It may be, for all we know, that our case of the soldiers obeying the Major is a trumping case that actually happens.

Fourth. There is another kind of causal connection to which the intuition that causation is an intrinsic matter does not apply. This is *double prevention*: a cause prevents something which, had it not been prevented, would have prevented the

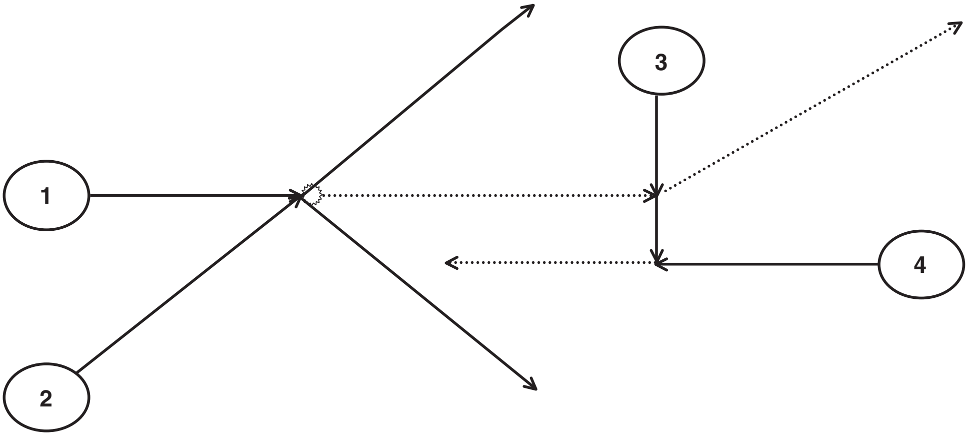

Figure 3.1

effect. The collision between billiard balls 1 and 2 prevents ball 1 from continuing on its way and hitting ball 3 (fig. 3.1). The collision of 1 and 3, had it occurred, would have prevented the subsequent collision of balls 3 and 4. But since in fact the collision of 1 and 3 was prevented, the collision of 3 and 4 was *un*prevented. That is how the collision of 1 and 2 causes the collision of 3 and 4. It's a straightforward case of counterfactual dependence: Without the collision of 1 and 2, the collision of 3 and 4 would not have occurred. But notice that this counterfactual dependence is an extrinsic matter. Had there been some other obstruction that would have stopped ball 1 from hitting ball 3, the collision of 3 and 4 would not have depended on the collision of 1 and 2. So even in this very ordinary thisworldly case, the causal connection is extrinsic.

Two more examples. Michael McDermott's: A crazed American President is about to launch a nuclear attack on Russia; that attack would have provoked a counterstrike, which would have prevented Joe Blow from eating breakfast the following day. Luckily, the President's assistant intervenes to stop the attack. Joe Blow's breakfast depends counterfactually upon that intervention. But the dependence is an extrinsic matter: Had Russia been uninhabited or unarmed, there would have been no such dependence (McDermott 1995a).

Ned Hall's: Billy, the pilot of the escort fighter, shoots down the interceptor that would otherwise have shot down the bomber. Therefore the successful bombing of the target depends counterfactually on Billy's action. But again the dependence is extrinsic. If the interceptor had been about to receive a radio order to return to base without attacking the bomber, then the successful bombing would not have depended on Billy's action.[12]

Fifth. Besides the overhasty intuition of the intrinsic character of causation, there is another presupposition of the method of quasi-dependence that also breaks down in cases of double prevention. That is the presupposition that we have a chain of events that runs from the preempting cause to the effect. We need this chain of events so that we can say what chain of events in the comparison case is its intrinsic duplicate. But when we have causation by double prevention, there is often no continuous chain of events running from cause to effect. Between the collision of balls 1 and 2 and the collision of balls 3 and 4, or between the intervention by the President's assistant and Joe Blow's breakfast, or between the shooting down of the interceptor and the bombing of the target, nothing much happens. What matters, of course, is what *doesn't* happen. Sometimes maybe we can assign definite locations to the prevented intermediates, and thereby locate a chain of events *and absences*. Sometimes not. If a preempting cause happens to work by double prevention—and, once we watch for them, cases of double prevention turn out to be very common—and if we cannot assign any definite location to the relevant absences, there is no saying what the intrinsic character of the comparison chain is required to match.

Put another way, the method of quasi-dependence breaks down when we have causation at a distance; and causation at a distance, rather than being the far-fetched possibility we might have thought it was, turns out to be a feature of commonplace cases of double prevention. What *is* far-fetched—though it may nevertheless turn out to be the truth about the collapse of spatially spread-out wave functions—is *action* at a distance; and that is only one variety of causation at a distance.[13,14] If, for instance, one body exerted a force on a distant body without any field or any particle going from one to the other, *that* would be action at a distance. Our billiard-table example of double prevention, however, exhibits quite a different kind of causation at a distance.

### 1.7 Fragility Corrected

There is an obvious solution to cases of late preemption. Doubtless you have been waiting impatiently for it. Without Suzy's preempting rock, the bottle would still have shattered, thanks to Billy's preempted rock. But this would have been a *different* shattering. It would, for instance, have happened a little later. The effect that actually occurred *did* depend on Suzy's throw. It did not likewise depend on Billy's. Sometimes this solution is just right, and nothing more need be said. Suppose it were alleged that since we are all mortal, there is no such thing as a cause of death. Without the hanging that allegedly caused Ned Kelly's death, for instance, he would sooner or later have died anyway. Yes. But he would have died a different death. The event which actually was Kelly's death would never have occurred.

The case of Suzy's preempting throw is different, however. It's not just that without it the bottle would have shattered somehow, sooner or later. Without it, the bottle would have shattered at very nearly the same time that it actually did shatter, in very nearly the same way that it actually did. Yet we're usually quite happy to say that an event might have been slightly delayed, and that it might have differed somewhat in this or that one of its contingent aspects. I recently postponed a seminar talk from October to December, doubtless making quite a lot of difference to the course of the discussion. But I postponed it instead of canceling it because I wanted *that very event* to take place.

So if we say that the shattering of the bottle was caused by Suzy's throw, because without it that very shattering would not have occurred, we are evoking uncommonly stringent conditions of occurrence for that event. We are thinking that it would take only a very slight difference to destroy that event altogether, and put a different substitute event in its place. We are supposing the shattering to be modally *fragile*. This is not something we would normally suppose. We have no business first saying as usual that the very same event might have been significantly delayed and changed, and then turning around and saying that it is caused by an event without which it would have been ever so slightly delayed and changed, and then saying that this is because it takes only a very slight delay or change to turn it into a different event altogether.

How much delay or change (or hastening) *do* we think it takes to replace an event by an altogether different event, and not just by a different version of the same event? An urgent question, if we want to analyze causation in terms of the dependence of whether one event occurs on whether another event occurs. Yet once we attend to the question, we surely see that it has no determinate answer. We just haven't made up our minds; and if we speak in a way that presupposes sometimes one answer and sometimes another, we are entirely within our linguistic rights. This is itself a big problem for a counterfactual analysis of causation, quite apart from the problem of preemption.[15]

At least, it is a problem so long as we focus on whether–whether counterfactual dependence. But there are other kinds of dependence. There is, for instance, when-on-whether dependence: When one event occurs depends counterfactually on whether another event occurs. And that is only the beginning. But even this beginning is enough to rehabilitate the obvious solution to late preemption, at least in very many commonplace cases. Let us by all means agree that Suzy's throw caused the shattering of the bottle because, without her throw, the shattering would have been slightly delayed. But let us not go on to say that if it had been slightly delayed, that would have turned it into a different event altogether. Let us rather say that Suzy's throw caused the shattering of the bottle in virtue of when-on-whether dependence. When the shattering occurred depended on Suzy's throw. Without Suzy's throw, it would not have occurred exactly when it actually did occur.

L. A. Paul has proposed an emended analysis of causal dependence: event *E* depends causally on a distinct actual event *C* if and only if "if *C* had not occurred, then *E* would not have occurred at all *or would have occurred later than the time that it actually did occur*" (Paul 1998b).[16] (Causation itself is the ancestral: *C* causes *E* iff there is a chain of such dependencies running from *C* to *E*.) This proposal does not abandon the strategy of fragility, but corrects it. Instead of supposing that the event itself is fragile—which would fly in the face of much of our ordinary talk—we instead take a tailor-made fragile proposition about that event and its time. The negation of that fragile proposition is the consequent of our causal counterfactual. Now we get the right answer to commonplace cases of late preemption. Suzy's throw hastens the shattering, Billy's doesn't. So Suzy's throw causes the shattering, Billy's doesn't.

If we stopped here, we would be building into our analysis an asymmetry between hasteners and delayers. We would be saying that an event without which the same effect would have happened later is a cause, whereas an event without which the same effect would have happened earlier is not.[17] For that reason, among others, we should not stop here. We should admit delayers as causes, even when the delayed event is the very same event that would otherwise have happened earlier—or at least, to acknowledge our indecision about such questions, not clearly *not* the same event.

We're often ambivalent about the status of delayers. Perhaps that is because a delayer often works by double prevention. It causes a later version of the event by preventing an earlier version, which, had it happened, would have prevented the later version. Then if we ask whether the delayer prevented the event or caused it, and we overlook the possibility that it might have done both, we have to say "prevented" (see Mackie 1992). To restore symmetry between hastening and delaying, we need only replace the words "or would have occurred later than the time that it actually did occur" by the words "or would have occurred at a time different from the time that it actually did occur." I favor this further emendation. (As does Paul.) But I think we should go further still. What's so special about time? When we thought that without the actual causes of his death, Ned Kelly would have died a different death, we were thinking not just that he would have died at a different time, but also that he would have died in a different manner. According to the uncorrected strategy of fragility, which supposes that events have very stringent conditions of occurrence, a difference either of time or of manner would suffice to turn the effect into a numerically different event. And if, imitating Paul's correction, we relocate the fragility not in the event itself but rather in a tailor-made proposition about that event, that will be a proposition about whether and when and how the effect occurs. We could further emend our analysis to require dependence of whether and when and how upon whether: Without *C*, *E* would not have occurred at all, or would have occurred at a time different from the time that it actually did occur, or would have occurred in a manner different from the manner in which it actually did occur. (And we could redefine causation as the ancestral of this new kind of dependence.)

This formulation still distinguishes the case that event *E* occurs differently from the case that *E* does not occur at all. The distinction has been made to not matter, but we're still presupposing that there is a distinction. If we're as indecisive about such questions as I think we are, it would be better to avoid that presupposition.

Let an *alteration* of event *E* be either a very fragile version of *E* or else a very fragile alternative event which may be similar to *E*, but is numerically different from *E*. One alteration of *E* is the very fragile version that actually occurs: the *unaltered* alteration, so to speak. The rest are unactualized. If you think *E* is itself very fragile, you will think that all its unactualized alterations are alternatives, numerically different from *E* itself. If you think *E* is not at all fragile, you will think that all its alterations are different versions of one and the same event. Or you might think that some are alternatives and others are versions. Or you might refuse to have any opinion one way or the other, and that is the policy I favor. Now we may restate our current analysis of causal dependence. We can return to whether–whether counterfactual dependence, but with alterations of the effect put in place of the event itself: Without *C*, the alteration of *E* which actually did occur would not have occurred. However indecisive we may be about how fragile an event itself is, its actual alteration is by definition fragile.

Now we say that Suzy's throw caused the shattering of the bottle and Billy's preempted throw did not because, without Suzy's throw, the alteration of the shattering which actually did occur would not have occurred, and a different alteration would have occurred instead. And here we are considering not only the slight delay before Billy's rock arrived but also any differences to the shattering that might have been made because Billy's rock differs from Suzy's in its mass, its shape, its velocity, its spin, and its aim point.[18]

## 1.8 Spurious Causation

We have dealt with one objection against the fragility strategy: that it conflicts with what we normally think about the conditions of occurrence of events. But there is a second objection, and it applies as much to the corrected strategy as to the strategy in its original form. All manner of irrelevant things that we would not ordinarily count among the causes of the effect can be expected to make some slight difference to its time and manner. I once gave this example: If poison enters the bloodstream more slowly when taken on a full stomach, then the victim's death, taken to be fragile—we might better say, the actual alteration of the victim's death—depends not only on the poison but also on his dinner.[19] If we heed still smaller differences, almost everything that precedes an event will be counted among its causes. By the law of universal gravitation, a distant planet makes some minute difference to the trajectory of Suzy's rock, thereby making a tiny difference to the shattering of the bottle. So by adopting the fragility strategy, whether in corrected or uncorrected form, we open the gate to a flood of spurious causes.

Among the spurious causes that should have been deemed irrelevant is Billy's rock, the preempted alternative. For one thing, it too exerts a minute gravitational force on Suzy's rock. We wanted to say that (the actual alteration of) the shattering depended on Suzy's throw and not on Billy's, but that turns out to be not quite true.

Well—these differences made by spurious causes are negligible, so surely we are entitled to neglect them? Just as it's right to say that a box contains nothing when, strictly speaking, it contains a little dust, so likewise we are within our linguistic rights to say that Billy's throw made no difference to the shattering when, strictly speaking, its gravitational effects made an imperceptibly minute difference. And if for some strange reason we chose to attend to these negligible differences, would we not then put ourselves in an unusual context where it is right, not wrong, to count all the things that make negligible differences as joint causes of the effect?

That would be a sufficient reply, I think, but for the fact that sometimes the difference made by a preempting cause is also minute. Imagine that Suzy's throw precedes Billy's by only a very short time; and that the masses, shapes, velocities, spins, and aim points of the two rocks also differ very little. Then without Suzy's throw we might have had a difference equal to, or even less than, some of the differences made by causes we want to dismiss as spurious.

But even so, and even if Billy's rock makes a minute difference to the shattering by way of its gravitational effects on Suzy's rock, yet Suzy's throw may make much *more* of a difference to the effect than Billy's. The alteration that would have occurred without Suzy's throw, though not very different from the actual alteration, may differ from it in time and manner much more than the alteration that would have occurred without Billy's. Though the difference made by Billy and the difference made by Suzy may both count as small by absolute standards, yet the difference made by Billy may be small also in comparison to the difference made by Suzy. That would be enough to break the symmetry between Suzy and Billy, and to account for our judgment that Suzy's throw and not Billy's causes the shattering. We speak of the asymmetry as if it were all-or-nothing, when really it is a big difference of degree, but surely such linguistic laxity is as commonplace as it is blameless.

If, on the other hand, Billy's throw does somehow make roughly as much difference to the effect as Suzy's, that is a good reason to judge that Billy's throw is not after all a mere preempted alternative. Rather it is a joint cause of the shattering. So in this case too we get the right answer.

## 2 Lecture II

### 2.1 Alterations of the Cause

Because we're so indecisive about the distinction between alterations that are different versions of the very same event and alterations that are different but similar events, we ought to make sure that this distinction bears no weight in our analyses. So far, we're obeying that maxim only one-sidedly. The distinction doesn't matter when applied to the effect, but it still matters when applied to the cause. What it means to suppose counterfactually that *C* does not occur depends on where we draw the line between *C* not occurring at all and *C* occurring differently in time and manner.

That makes a problem. What is the closest way to actuality for *C* not to occur?—It is for *C* to be replaced by a very similar event, one which is almost but not quite *C*, one that is just barely over the border that divides versions of *C* itself from its nearest alternatives. But if *C* is taken to be fairly fragile, then if *C* had not occurred and almost-*C* had occurred instead, very likely the effects of almost-*C* would have been much the same as the actual effects of *C*. So our causal counterfactual will not mean what we thought it meant, and it may well not have the truth value we thought it had.[20] When asked to suppose counterfactually that *C* does not occur, we don't really look for the very closest possible world where *C*'s conditions of occurrence are not quite satisfied. Rather, we imagine that *C* is completely and cleanly excised from history, leaving behind no fragment or approximation of itself. One repair would be to rewrite our counterfactual analysis, or add a gloss on its interpretation, in order to make this explicit (Lewis 1986b, p. 211).

But there is another remedy. We could look at a range of alterations of *C*, not just one. As on the side of effects, we need not ever say which of these are versions of *C* and which if any are alternatives to *C*. These alterations may include some in which *C* is completely excised, but we need not require this. They may include some which are almost but not quite $C$, but nothing is to be said that restricts us to the closest possible alterations. Then we look at the pattern of counterfactual dependence of alterations of the effect upon alterations of the cause. Where $C$ and $E$ are distinct actual events, let us say that *C influences E* iff there is a substantial range $C_1, C_2, \ldots$ of different not-too-distant alterations of $C$ (including the actual alteration of $C$) and there is a range $E_1, E_2, \ldots$ of alterations of $E$, at least some of which differ, such that if $C_1$ had occurred, $E_1$ would have occurred, and if $C_2$ had occurred, $E_2$ would have occurred, and so on. Thus we have a pattern of counterfactual dependence of whether, when, and how on whether, when, and how. (As before, causation is the ancestral: $C$ causes $E$ iff there is a chain of stepwise influence from $C$ to $E$.) Think of influence this way. First, you come upon a complicated machine, and you want to find out which bits are connected to which others. So you wiggle first one bit and then another, and each time you see what else wiggles. Next, you come upon a complicated arrangement of events in space and time. You can't wiggle an event: it is where it is in space and time, there's nothing you can do about that. But if you had an oracle to tell you which counterfactuals were true, you could in a sense "wiggle" the events; it's just that you have different counterfactual situations rather than different successive actual locations. But again, seeing what else "wiggles" when you "wiggle" one or another event tells you which ones are causally connected to which.

A process capable of transmitting a mark, in the sense of Reichenbach and Salmon, is a good example of influence (Reichenbach 1928, sections 21 and 43; Salmon 1994). We have some sort of process extending along a continuous spatiotemporal path. We can mark the process at one stage, and that mark will persist at later stages. Or rather—since it is irrelevant whether there is actually anything around that can make a mark—if the process *were* somehow marked at one stage, that mark *would* persist at later stages. That is, we have patterns of influence whereby alterations of later stages depend counterfactually on alterations of earlier stages.[21] The process capable of transmitting a mark might, for instance, be a flow of energy, matter, momentum, or some other conserved quantity: if there were a little more or less of the quantity at an early stage, there would be correspondingly more or less of it at later stages (Fair 1979; Dowe 1992).

But transmission of a mark is only one special case of a pattern of influence. In general, we do not require that the alterations of $E$ resemble the alterations of $C$ that map onto them. Nor do we require that sufficiently similar alterations of $C$ map onto similar alterations of $E$. Nor do we require a process along a spatiotemporally continuous path; we could have influence of $C$ upon $E$ even if these were two separated events with nothing relevant between them. And we do not require a many–many mapping; the simplest sort of whether–whether dependence, with only two different alterations of $E$, still qualifies as one sort of pattern of influence. Recall Hall's example of causation by double prevention: The shooting down of the interceptor causes the destruction of the target by preventing the shooting down of the bomber. The shooting down of the interceptor does not much resemble the destruction of the target; there was no continuous process linking cause and effect; and alterations of the cause would in some cases have prevented the effect and in some cases not, but in no case would they have made a (more than negligible) difference to the effect without preventing it altogether.[22]

Influence admits of degree in a rough and multidimensional way. How many different $C_i$s are there? How distant are the rest of them from the actual alteration of $C$, and from one another? How much do the $E_i$s differ from one another: How many different ones are there, and when two of them do differ, how distant (on average, or at maximum) are they? Plainly there are many ways in which something can be more of a cause of some effect than something else is, even if it is not an all-or-nothing difference of influence versus no influence.

Now we are in a better position than before to say that Suzy's throw is much more of a cause of the bottle's shattering than Billy's. Even if the throws are so much alike that removing Suzy's throw altogether would make little difference to the shattering, it's still true that altering Suzy's throw while holding Billy's fixed would make a lot of difference to the shattering, whereas altering Billy's throw while holding Suzy's fixed would not. Take an alteration in which Suzy's rock is heavier, or she throws a little sooner, or she aims at the neck of the bottle instead of the side. The shattering changes correspondingly. Make just the same alterations to Billy's preempted throw, and the shattering is (near enough) unchanged.[23]

(Although Billy's throw does not influence the shattering, Billy's *not* throwing before the time of Suzy's throw does. This is a typical example of a delaying cause. The doctors who treated Ned Kelly's wounds lest he cheat the hangman by dying prematurely were the hangman's accomplices: joint causes, along with the judge and the hangman, of the death Ned actually died. Likewise Billy's earlier nonthrow and Suzy's throw were joint causes of the shattering that actually occurred.)

Thanks to this latest emendation of the counterfactual analysis, cases of trumping are covered along with commonplace preemption. Sergeant and Major both shout "Advance!" The soldiers advance. Altering the Major's command while holding the Sergeant's fixed, the soldiers' response would have been correspondingly altered. If the Major had said "Take cover!" they would have taken cover, if he had said "Retreat!" they would have retreated, and so on. Altering the Sergeant's command while holding the Major's fixed, on the other hand, would have made (near enough) no difference at all. If we look only at the whether–whether dependence of the soldiers' response on the actual commands of the two officers, we miss exactly the sort of counterfactual dependence that breaks the symmetry between the two.[24]

Likewise for the two wizards. If Merlin's first spell of the day had been not prince-to-frog, but rather king-to-kangaroo, the transmogrification at midnight would have been correspondingly altered. Whereas if Morgana's trumped spell had been, say, queen-to-goanna (holding fixed Merlin's earlier spell and the absence of any still earlier spell) what happened at midnight would have been exactly the same as it actually was: The prince would have turned into a frog, and that would have been all.

(Simon Keller has objected as follows. Suppose the soldiers are not perfectly obedient, and they know that the Sergeant is better placed than the Major to spot approaching danger. The Sergeant and the Major both shout "Retreat!" The soldiers infer from the Sergeant's order that they are in danger, so they retreat. The Sergeant's order causes the retreat. Yet if the Sergeant's order had been anything else, they would not have inferred danger, so they would have obeyed the Major.—Reply: the soldiers think "This is one of those exceptional times when it's best to obey the Sergeant." There is a range of alterations of the Sergeant's order, namely the range of alterations in which this thought is held fixed, for which we would have corresponding alterations of the soldiers' response. True, if the Sergeant's order had been different, this thought would not have been there. But even when it's true that if $P$, it would not have been that $Q$, we can still entertain the counterfactual supposition that $P$ and $Q$.[25] And we have not restricted ourselves to the alterations that are closest to actuality.)

## 2.2 Transitivity of Causation

Causation, I previously said, is the ancestral of causal dependence. Event $C$ causes event $E$ iff there is a chain of dependencies running from $C$ to $E$. That part of my analysis has remained untouched, even as my definition of causal dependence evolved from simple whether–whether dependence between events to a pattern of influence. Is it still necessary to take the ancestral? Or does our improved definition of causal dependence as a pattern of influence allow us just to identify causation with dependence?—No. Influence is not invariably transitive. If we want to ensure that causation is invariably transitive, we still have to take an ancestral.

You might think that intransitivities of influence could arise from intransitivities of the counterfactual conditional itself. We know that it can be true that if $P$, it would be that $Q$, and true also that if $Q$, it would be that $R$, yet false that if $P$, it would be that $R$ (see my 1973b, pp. 32–33; Stalnaker 1968). But that is not the problem. Though counterfactual transitivity itself is fallacious, a closely related inference

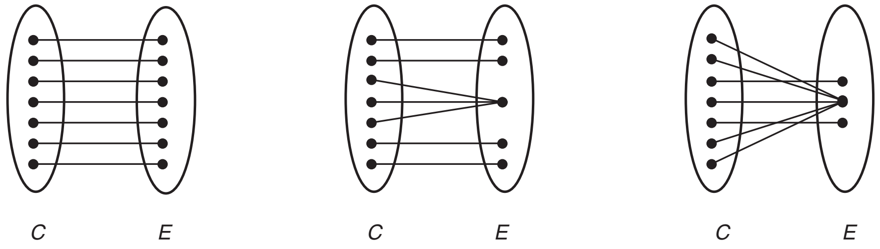

**Figure 3.2**

pattern is valid: from the premise that if *P*, it would be that *Q*, and the premise that if *both P* and *Q*, it would be that *R*, it does follow that if *P*, it would be that *R* (Lewis 1973b, p. 35). Let the counterfactual from $C_i$ to $D_i$ be part of a pattern of influence of *C* on *D*; let the counterfactual from $D_i$ to $E_i$ be part of a pattern of influence of *D* on *E*; then it would seem that if both $C_i$ and $D_i$, it would be that $E_i$; so we do indeed have the counterfactual from $C_i$ to $E_i$, and likewise for the other counterfactuals that constitute a pattern of influence of *C* on *E*.

The real problem with transitivity is that a pattern of influence need not map *all* the not-too-distant alterations of *C* onto different alterations of *D*, or all the not-too-distant alterations of *D* onto different alterations of *E*. Transitivity of influence can fail because of a mismatch between the two patterns of influence.

In figure 3.2 I picture three possible patterns of influence of *C* on *E*. The first is nice and simple: it maps several alterations of *C* one–one onto alterations of *E*. But less nice patterns will still qualify. Let the actual alteration be at the center, and imagine that distance from the center somehow measures closeness to actuality. (There's no need to make this distinction of inner and outer precise. Its only point is to make the cases easier to picture.) We might have a pattern of influence that maps the outer alterations of *C* one–one onto different alterations of *E*, but funnels all the inner alterations alike onto a single point (second picture). Or we might have a pattern that maps the inner alterations of *C* one–one onto different alterations of *E*, but funnels all the outer alterations alike onto a single point (third picture).

Now suppose *C* influences *D* by a pattern that funnels all the inner alterations onto a single point, while *D* influences *E* by a pattern that funnels all the outer alterations onto a single point (leftmost picture in fig. 3.3); or vice versa (middle picture). Or we might have more complicated cases (rightmost picture). In each case, the two patterns of influence that take us from *C* to *D* to *E* are mismatched: The values of the first pattern do not coincide with the arguments of the second. So *C* influences *D* and *D* influences *E*, but *C* does not influence *E*. If we nevertheless want to say

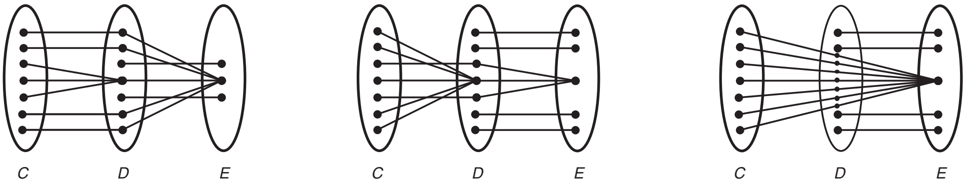

**Figure 3.3**

that *C* causes *E*, we have to take the ancestral and say that causation outruns direct influence.

How might such a case arise? Here is a famous example (from Frankfurt 1969; see also Heinlein 1951). The Neuroscientist knows exactly how she wants Jones to behave. She hopes that Jones, left to himself, will behave just as she wants him to. By reading his brain, she can predict what he will do if left to himself. She reads that he will do what she wants, so she does nothing further. But if instead she had read that he would stray from the desired path, she would have taken control. She would have made him a puppet, manipulating his brain and nervous system directly so as to produce the desired behavior. The initial state of Jones's brain is a preempting cause of his behavior; the idle Neuroscientist is a preempted backup. The moral of the story is that preemptive causation, without dependence, suffices to confer ownership and responsibility for one's actions.

Let *C* be Jones's initial brain state; let *E* be the desired behavior. Consider a time after the Neuroscientist has read Jones's brain, but before she would have seized control if the reading had been different. Let *D* combine Jones's brain state at that time with the Neuroscientist's decision not to intervene. *C* influences *D*. *D* in turn influences *E*, since at the time of *D* it's too late for the Neuroscientist to intervene. So we have a two-step chain of influence from *C* to *D* to *E*. But *C* does not influence *E*: any alteration of Jones's initial brain state would have led to the same behavior in the end, one way or the other.

The actual alteration of *C* is the one (assume it to be unique) that leads to exactly the desired behavior. The actual alteration of *E* consists of the desired behavior; the other alterations of *E* consist of different behavior. The actual alteration of *D* is the one that leads to the desired behavior, and that includes the Neuroscientist's decision not to intervene. The "inner" alterations of *D* are those that would not lead to the desired behavior, but that include the Neuroscientist's decision to intervene in one or another way. The "outer" alterations of *D* are those that would not lead to the desired behavior, but that nevertheless include the Neuroscientist's decision not to intervene.[26] These are arguments of the pattern of influence from *D* to *E*, and without them it would not be a pattern of influence at all. But they are not among the values of the pattern from *C* to *D*. The pattern of influence of *C* on *D* maps the actual alteration of *C* onto the actual alteration of *D*, and all other alterations of *C* onto inner alterations of *D*. The pattern of influence of *D* upon *E* maps all the inner alterations of *D* onto the actual alteration of *E*, and the outer alterations of *D* onto different alterations of *E*. Feeding the first pattern into the second, we get a pattern which maps all alterations of *C* onto the actual alteration of *E*. Thus the patterns are mismatched in the way shown in the rightmost picture in figure 3.3. Transitivity of influence fails.

This is an easy case of early preemption—just the sort of case that my strategy of taking the ancestral was originally made for. If we'd tried to make do without the ancestral, and get by with influence alone, it would remain unsolved—provided that we insist, as of course we should, that, with no intervention at all by the Neuroscientist, Jones's initial brain state is indeed a cause of his behavior.

### 2.3 Transitivity Defended

Some will say that by making causation invariably transitive, our strategy of taking the ancestral makes more trouble than it cures. It collides with a flock of alleged counterexamples to transitivity of causation. Thus I've incurred an obligation to deal with these examples.

The alleged counterexamples have a common structure, as follows. Imagine a conflict between Black and Red. (It may be a conflict between human adversaries, or between nations, or between gods striving for one or another outcome, or just between those forces of nature that conduce to one outcome versus those that conduce to another.) Black makes a move that, if not countered, would have advanced his cause. Red responds with an effective countermove, which gives Red the victory. Black's move causes Red's countermove, Red's countermove causes Red's victory. But does Black's move cause Red's victory? Sometimes it seems not.

One of the best known of these Black–Red counterexamples comes from Jonathan Bennett (1987). *Forest fire*: Let Black be those forces of nature that want the forest to survive; let Red be those forces of nature that want it to burn. Black protects the forest from the May lightning by raining all over it in April. Red dries the forest off again before more lightning comes. The forest burns in June. The April rain caused there to be an unburnt forest in June, which in turn caused the June fire. If causation is invariably transitive, we must conclude that the rain caused the fire.

Two more come from Michael McDermott (1995a). *Shock C*: Black is *C*'s friend, Red is *C*'s foe. *C* will be shocked iff the two switches are thrown alike. Black, seeing that Red's switch is initially thrown to the left, throws his switch to the right. Red, seeing this, responds by throwing his switch also to the right. $C$ is shocked. Black's throwing his switch caused Red to throw his switch, which in turn caused $C$ to be shocked. Thus Black's attempt to protect $C$ is thwarted. If causation is invariably transitive, Black's failed attempt to prevent the shock is actually among the causes of the shock.

*Dog-bite*: Red wants to cause an explosion; Black (nature) wants him not to. Black's move: a dog bites off right-handed Red's right forefinger. Red's countermove: with difficulty, he uses his left hand to set off the bomb. The bomb explodes. The dog-bite caused Red to set off the bomb with his left hand, which in turn caused the explosion. If causation is invariably transitive, the dog-bite was a cause of the explosion.

Another comes from Hartry Field (unpublished lecture). *The bomb outside the door*: Black wants Red dead, so he leaves a bomb outside Red's door. Red finds it and snuffs out the fuse. Red survives. Placing the bomb caused Red to snuff out the fuse, which in turn caused Red's survival. If causation is invariably transitive, placing the bomb was a cause of Red's survival.

Three more examples come from Ned Hall ("Two Concepts of Causation"). *The deadly double dose*: Black endangers Billy by giving him half of the deadly double dose on Monday. Red counters by withholding the second half on Tuesday. Billy survives. Monday's dose caused Tuesday's withholding, which in turn caused Billy's survival. If causation is invariably transitive, Monday's dose was a cause of Billy's survival.

*The alarm clock*: The ringing of the alarm clock summons the Black champion forth into battle, where he is slain by the Red forces. Without him, Black's cause is lost. Red wins. The ringing clock caused the champion to be slain, which in turn caused Red's victory. If causation is invariably transitive, the ringing clock was a cause of Red's victory.

*The inert network* (fig. 3.4): Red wants neuron $F$ to fire, Black wants it not to. Since $F$ is extraneously stimulated, it will fire unless it is somehow inhibited. Black's move: Fire $C$, which has a stimulatory connection to $D$, which in turn has a stimulatory connection to $E$, which in turn has an inhibitory connection to $F$. Red's countermove (made in advance): Provide another stimulatory connection from $C$ to $B$, which in turn has an inhibitory connection to $E$. So $E$ doesn't fire, $F$ is uninhibited, and $F$ does fire. The neural network consisting of $C, D, B$, and $E$ is inert, so far as $F$ is concerned; there's no way it could have prevented $F$ from firing. Yet the firing of $C$ caused the firing of $B$; which in turn caused the *non*firing of $E$, which in turn caused the firing of $F$. If causation (including causation by double prevention) is invariably transitive, then the firing of $C$ was a cause of the firing of $F$.

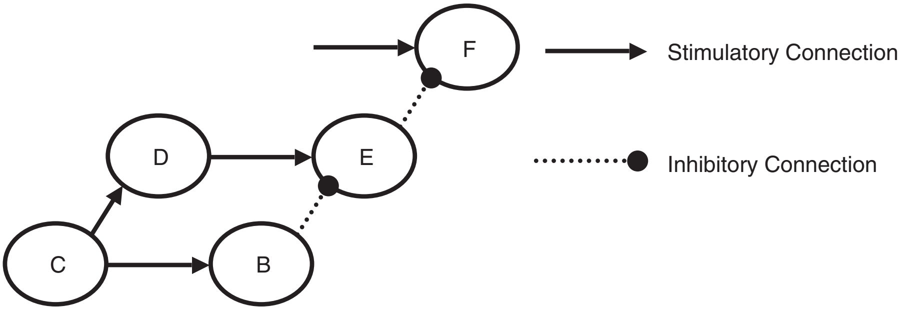

**Figure 3.4**

My last example is suggested by a familiar saying, *Damned if you do, damned if you don't*: Black tries to do what God has commanded, but the Red Devil interferes so that he messes it up. There ain't no justice: God accepts no excuses. So Black is damned. Black's failed attempt at pious obedience caused the Devil to interfere, which in turn caused Black to be damned. If causation is invariably transitive, Black's pious conduct caused him to be damned. In all these cases, there are two causal paths the world could follow, both leading to victory for Red. The two paths don't quite converge: Victory may come in one way or in another, it may come sooner or it may come later, but Red wins in the end. Black's thwarted attempt to prevent Red's victory is the switch that steers the world onto one path rather than the other. That is to say, it is because of Black's move that Red's victory is caused one way rather than the other. That means, I submit, that in each of these cases, Black's move does indeed cause Red's victory. Transitivity succeeds.

That is my considered opinion, but I do admit to feeling some ambivalence. Insofar as I can conjure up any inclination to accept the counterexamples, I think my inclination has three sources, all of them misguided.[27]

First. In many of these cases Red's victory would have come sooner, or more directly, without Black's move. Black's move prevents Red's victory as well as causing it: It causes one version but it prevents another. If we thought we had to choose, we would wrongly infer that since it is a preventer it cannot be a cause. (We've already noted this ambivalence in the case of delaying causes generally.)

Second. Moves such as Black's are in general conducive to victory for Black, not for Red. If we mix up questions of what is generally conducive to what with questions of what caused what in this particular case, we may think it just a bit of good common sense to say that Black's moves advance Black's cause, not Red's.[28]

Third. We note that Black's move didn't matter; Red would have won all the more easily without it. The effect doesn't depend on the cause. The idea that causation requires whether–whether dependence may retain some grip on us. But if you *ever* accept preemptive causation, you must have learned to resist that grip. Why yield to it now? It's true that Black's move didn't matter. But that's because the choice Black faced (whether he knew it or not) was whether to have his defeat caused in one way or in another; and, either way, Black's defeat is *caused*.

In rejecting the counterexamples, and accepting that Black's move is a cause of Red's victory, I think I am doing what historians do. They trace causal chains, and, without more ado, they conclude that what comes at the end of the chain was caused by what went before. If they did not, they could say little about historical causation; because, over intervals of any length, historical counterfactuals become so very speculative that nothing much can be known about the dependence of any event on its causal ancestors. And every historian knows that actions often have unintended and unwanted consequences. It would be perfectly ordinary for a move by Black to backfire disastrously.

I've assumed so far that the Black–Red examples are genuine test cases: We really do have an event $C$ that causes an event $D$ that in turn causes an event $E$. But unless the examples are carefully formulated, perhaps with the aid of somewhat artificial stipulations, that may not be so. It may rather be that $C$ causes $D_1$ and $D_2$ causes $E$; and $D_1$ and $D_2$ are different, even though perhaps we may refer to them by the same name. If so, the example is not a test case, and if it turns out (contrary to my opinion) that $C$ does not cause $E$, that is no problem for the thesis that causation is invariably transitive.

$D_1$ and $D_2$ might, for instance, be two different aspects of the same event: $D$-*qua*-event-of-kind-$A$ and $D$-*qua*-event-of-kind-$B$ (see Paul, "Aspect Causation," chapter 8 in this volume). Or $D_1$ and $D_2$ might be $D$ taken with two different contrasts: $D$-rather-than-$X$ and $D$-rather-than-$Y$ (see Maslen, "The Context-Dependence of Causation," chapter 14 in this volume; Hitchcock 1996b). The contrast might be supplied tacitly by contextual clues, or it might be explicit. I think the aspect proposal and the contrast proposal don't differ much: The aspect $D$-*qua*-event-of-kind-$A$ pretty much amounts to the contrasted event $D$-rather-than-a-version-of-$D$-that-is-not-of-kind-$A$. I'd suggest that aspects and contrasts alike are best understood as constraints on the range of relevant alterations.

## 2.4 Causation by Absences

Alterations, I said, are very fragile events. That was not quite right: Some of them are absences. Absences can be causes, as when an absence of food causes hunger.

Absences can be effects, as when a vaccination prevents one from catching a disease. And absences can be among the unactualized alterations of a cause or effect that figure in a pattern of influence.

Absences are not events. They are not *anything*: Where an absence is, there is nothing relevant there at all.[29] Absences are bogus entities. Yet the proposition that an absence occurs is not bogus. It is a perfectly good negative existential proposition. And it is by way of just such propositions, and only by way of such propositions, that absences enter into patterns of counterfactual dependence. Therefore it is safe to say with the vulgar that there are such entities as absences, even though we know better. If there is no more beer in the fridge, it is a fiction that the beer has been replaced by something else, something called an "absence of beer." We can say that there's an absence of beer, sure enough; and it's part of the fiction that this proposition is made true by the existence of the absence. But the sober truth is rather that this proposition is true because the proposition that there is some beer is false. That said, I also insist that the fiction is harmless, and we are within our linguistic rights to indulge in it. Accordingly, I shall carry on make-believedly quantifying over absences without apology.

(Should we conclude, then, that when we say that absences are causes, really it is true negative existential propositions that do the causing?—No; in other cases we distinguish between the cause itself and the true proposition that describes it. For instance, we distinguish the explosion from the proposition that an explosion occurred at so-and-so place and time. The explosion caused the damage; the proposition is a necessary being, "abstract" in one sense of that multifariously ambiguous term, and doesn't cause anything. On absences, as also on the aspects of events, I have met the friends of "fact causation" more than halfway; but I refuse to concede that facts—true propositions—are literally causes.[30] So I have to say that when an absence is a cause or an effect, there is strictly speaking nothing at all that is a cause or effect. Sometimes causation is not a relation, because a relation needs relata and sometimes the causal relata go missing [see my "Void and Object," chapter 10 in this volume]. But often, when one genuine event causes another, there are relata, and a causal relation that holds between them. So if we ignore all causal judgments except those framed by putting a "because" between clauses that express propositions, we overlook part of our topic.)

One reason for an aversion to causation by absences is that if there is any of it at all, there is a lot of it—far more of it than we would normally want to mention. At this very moment, we are being kept alive by an absence of nerve gas in the air we are breathing. The foe of causation by absences owes us an explanation of why we sometimes do say that an absence caused something. The friend of causation by absences owes us an explanation of why we sometimes refuse to say that an absence caused something, even when we have just the right pattern of dependence.[31] I think the friend is much better able to pay his debt than the foe is to pay his. There are ever so many reasons why it might be inappropriate to say something true. It might be irrelevant to the conversation, it might convey a false hint, it might be known already to all concerned, and so on (Grice 1975).

Of course, such reasons for refusing to say what's true are not confined to causation by absences. "Counterfactual analysis of causation?—Yeah, yeah, my birth is a cause of my death!" said the scoffer. His birth is indeed a cause of his death; but it's understandable that we seldom want to say so. The counterfactual dependence of his death on his birth is just too obvious to be worth mentioning.

(In case you're tempted to agree with the scoffer, consider this comparison of cases. In actuality there are no gods, or anyway none who pay any heed to the lives of mere mortals. You are born, and after a while you die. In the unactualized comparison case, the gods take a keen interest in human affairs. It has been foretold that the event of your death, if it occurs, will somehow have a momentous impact on the heavenly balance of power. It will advance the cause of Hermes, it will be a catastrophe for Apollo. Therefore Apollo orders one of his underlings, well ahead of time, to see to it that this disastrous event never occurs. The underling isn't sure that just changing the time and manner of your death would suffice to avert the catastrophe; and so decides to prevent your death altogether by preventing your birth. But the underling bungles the job: you are born, you die, and it's just as catastrophic for Apollo as had been foretold. When the hapless underling is had up on charges of negligence, surely it would be entirely appropriate for Apollo to complain that your birth caused your death. And if it's appropriate to say, presumably it must be true. But now we may suppose that, so far as earthly affairs go, actuality and our unactualized comparison case are alike in every detail. After all, the underling didn't manage to do anything. We may also suppose that, so far as earthly affairs go, the two cases are subject to exactly the same laws of nature. So, if you agree with the scoffer that your birth didn't cause your death in actuality, you must think that idle heavenly differences can make a difference to what causes what here below! That is hard to believe. To be sure, we earlier dismissed the thesis of the intrinsic character of causation as an overhasty generalization. But here, all we need is that earthly causal relations supervene on the intrinsic and nomological character of all things earthly.)

As I mentioned previously, Jaegwon Kim has drawn our attention to several causes of noncausal counterfactual dependence. I said in reply that counterfactual dependence is causal when it is dependence between entirely distinct events, neither identical nor overlapping; and that events (or at least, those of them that are causal relata) must be predominantly intrinsic (see Kim 1973b and my 1986d). Xanthippe's becoming a widow is a particular having of an extrinsic property; so it is not an event at all (or anyway, it is not a causal relatum), unless it is taken to be identical to, rather than distinct from, the event of Socrates' death.

When we say that absences as well as events can be causes and effects, do Kim's problems reappear? I think not. First, it is hard to see how an absence could be essentially a having of an extrinsic property. Second, it is safe to say that absences and genuine events are always distinct from one another. And third, we can say when two absences are distinct from one another: namely, when the corresponding negative existential propositions are logically independent.

It doesn't make sense for two distinct absences to differ slightly in detail. When we have an absence, there's nothing (relevant) there at all, and that's that. So when an absence is caused, we would expect a pattern of influence that exhibits funneling to an unusual degree. We can imagine a device that works in an extraordinarily precise all-or-nothing fashion; or a Neuroscientist, or some other marvelous being, able to exert extraordinarily precise and complete control; or we can just imagine a perfectly ordinary case of prevention. If we then follow that with the funneling that comes from the presence of a preempted backup, we may well end up with a mismatch between patterns of influence in which transitivity of influence fails. Small wonder, then, that cases of *preemptive prevention*—preemptive causing of an absence—and preemptive double prevention have appeared along with the Black–Red examples in the debate over transitivity of causation. I say again that at worst we have causation without direct influence. I trace a chain; I take the ancestral; I say that when a preempted preventer causes an absence which in turn causes some further event or absence, then the preempted preventer is a cause of that further event or absence.

Part of what makes preemptive prevention hard, however, is doubt about whether the absence really does cause anything further. Here is an example, due to Michael McDermott.[32] The fielder catches the ball; he causes its absence just beyond his hand. But a little further along its path there is a wall—a high, broad, thick, sturdy wall. Further along still is a window. Does the fielder cause the window to remain unbroken? Does he thereby cause the owner of the window to remain cheerful?

We are ambivalent. We can think: Yes—the fielder and the wall between them prevented the window from being broken; but the wall had nothing to do with it, since the ball never reached the wall; so it must have been the fielder. Or instead we can think: No—the wall kept the window safe regardless of what the fielder did or didn't do.

A treatment of the case ought to respect our ambivalence. Rather than endorsing the "Yes" or the "No," it ought to show how we are within our linguistic rights in giving either answer. The indeterminacy of our naive judgments is best explained by invoking some indeterminacy in our analysis. We are in a position to do this.

We have *C*, the catch. We have *D*, the absence of the ball from the place just beyond the fielder's hand. We have *E*, the absence of the impact of the ball on the window, or the nonbreaking of the window, or the continued good cheer of the owner. Certainly we have a pattern of influence of *C* on *D*. Whether we have influence of *D* on *E* is doubtful. There are alterations of *D* in which not only is the ball present beyond the fielder's hand, but also it is on a trajectory that would take it over the high wall and down again, or it is moving with energy enough to break through the wall, and so on. Some of these alterations of *D* would indeed have led to alterations of *E*. But are they relevant, "not-too-distant," alterations? We may be in a mood to think so, or we may be in a mood to think not. If we are in a mood to think them relevant, we should conclude that *D* causes *E*, and by transitivity *C* also causes *E*. That is the mood we are in when we are swayed by the thought that the fielder and the wall between them prevented the window from breaking. Whereas if we are in a mood to think them not relevant, we should conclude that neither *D* nor *C* causes *E*, and so the question of transitivity from *C* to *D* to *E* does not arise. That is the mood we are in when we are swayed by the thought that the window was safe regardless. But if anyone says that *D* causes *E* but *C* doesn't, and concludes that transitivity fails, he is not stably in one mood or the other.

The Yale shadow puzzle is similar. Two opaque objects are between the sun and the ground, in such a way that either one without the other would cast exactly the same shadow. (There might be more than two; and they might even be many slices of a single thick object.) The upper one is illuminated and stops the sunlight; the lower one is unilluminated. Does the upper one cast a shadow on the ground? We can think: Yes—between them, the two cast a shadow, but the lower one stops no light because no light ever reaches it, so the upper one must have done the job. Or we can think: No—thanks to the lower one, the ground would have been shadowed regardless of whether the upper one was there or not.[33] Again our ambivalence ought to be respected. We can explain it as before. Consider the absence of light just beyond the upper object; some far-fetched alterations of this absence would result in light getting through or around the second object, but we may well be of two minds about whether those alterations are too distant from actuality to be considered.

Yet another example of preemptive prevention comes from Ned Hall. The bomber is protected by two escort fighters, piloted respectively by Billy and Hillary. When the enemy interceptor arrives, Billy shoots it down; but had Billy failed, Hillary would have succeeded. In either case, the shooting down of the interceptor prevents the shooting down of the bomber, which, had it happened, would have prevented the subsequent bombing of the target (Hall, "Two Concepts of Causation," chapter 9 in this volume). What Hall says about this case parallels the thoughts that favored saying that the fielder prevented the window from breaking, or that the upper object cast the shadow on the ground: "If Billy's action was a cause of the bombing ... where Hillary was absent, then so too in this second case, which merely adds an alternative that plays no active role." Hall's view is defensible, provided he is in a mood not to ignore those far-fetched alterations in which the interceptor succeeds in evading both Billy and Hillary. But if so, then it is misleading (though literally true) for him to deny, as he does, that the bombing depends on Billy's action. Ignoring the far-fetched alterations, it's false that without Billy's action the bombing would have occurred anyway; what's true is that it might or might not have occurred. Saying that it would have occurred is equally defensible—but that calls for a different mood, one in which those far-fetched alterations *are* ignored.

## Notes

[1] See, *inter alia*, D. M. Armstrong, "Going through the Open Door Again," in this volume.

[2] Or perhaps I feel a pressure on my own body; and perhaps it is analytic that a pressure involves a force, and that a force is *inter alia* something apt for causing (or preventing) motion. Then it seems that I'm causally acquainted, if not with causation itself, at least with something conceptually linked to causation. See Armstrong, "The Open Door," this volume; and his (1962), p. 23, and (1997), pp. 211–216.

[3] The Canberra plan is derived, on one side, from Carnap's ideas about analyticity in a theoretical language; and on the other side, from one version of functionalism about mental states. See *inter alia* Carnap (1963), pp. 958–966; my (1966); and Armstrong (1968). The Canberra plan has been applied to causation in Tooley (1987), and in Menzies (1996). I discuss Menzies's treatment in "Void and Object," in this volume, saying that at best Menzies's approach will succeed in defining one central kind of causation.

[4] We could subsume the perceptual acquaintance strategy under the Canberra plan. We could take the job description that specifies the role occupied by causation to consist almost entirely of platitudes about how we are perceptually acquainted with causation.

[5] *An Enquiry Concerning Human Understanding*, section VII.

[6] For discussion of the problem, see *inter alia* Menzies (1989a); Woodward (1990); and Schaffer (2000a). Schaffer's version of the problem resists some of the strategies that might solve other versions.

[7] Two points of terminology. Some say "overdetermination" to cover all sorts of redundancy; I limit it to the symmetrical cases. Some say "preemption" to cover only those asymmetrical cases that do involve cutting; I apply it to all asymmetrical cases.

[8] For my former view, see the treatment of preemption in my (1986b), pp. 193–212. For a recent claim that all preemption involves cutting, see Ramachandran (1997a), p. 273: "... in all genuine causes of causal pre-emption, ... the pre-empted processes do not run their full course. ... All genuine causes, on the other hand, *do* seem to run their full course; indeed, they presumably count as genuine precisely because they do."

[9] This example, and the discovery of trumping, are due to Jonathan Schaffer. See his "Trumping Preemption," in this volume. The case of the soldiers is due to Bas van Fraassen.

[10] See my (1986b), pp. 205–207. For a similar proposal, see Menzies (1996) and (1999).

[11] Here I am pretty much following Paul (1998a).

[12] Hall (1994); Hall, "Two Concepts of Causation," chapter 9 in this volume.

[13] For a misguided conflation of causation at a distance with action at a distance, and consequent dismissal of causation at a distance as far-fetched, see my (1986b), p. 202.

[14] Perhaps action at a distance is "production" at a distance, where production is one of the varieties of causation distinguished by Ned Hall in "Two Concepts of Causation." Or perhaps it is "process-linkage" at a distance, where process linkage is explained as in Schaffer (2001).

[15] It is a problem that is seldom noted. However, see Bennett (1988), passim.

[16] "Actually" is right, strictly speaking, only if the causal connection in question is set in the actual world. More generally, $E$ depends causally on $C$ in world $W$ iff $C$ and $E$ occur in $W$ and it's true in $W$ that without $C$, $E$ would not have occurred or would have occurred later than it did in $W$. But we need not speak so strictly.

[17] For advocacy of just such an "asymmetry fact," see Bennett (1987); for reconsideration and rejection of it, see Bennett (1988), pp. 69–72.

[18] Here I've adopted a suggestion made by D. H. Rice at the Oxford Philosophical Society in 1984: "If $C_1$ and $C_2$ are redundant causes of $E$, and $E$ would have occurred more or less just as it did if $C_2$ had not occurred, but would not have occurred more or less just as it did if $C_1$ had not occurred, then $C_1$ is a cause simpliciter of $E$ and $C_2$ is not."

[19] Lewis (1986b), pp. 198–199. Here I am indebted to Ken Kress.

[20] See Bennett (1987), pp. 369–370. (Bennett's point here is independent of his defense of a hastener-delayer asymmetry elsewhere in that article.)

[21] Salmon abandoned his mark transmission account of causal processes after Nancy Cartwright and Philip Kitcher convinced him that it would need to be formulated in terms of counterfactuals; see Salmon (1994).

[22] However, some cases of causation by double prevention do exhibit a many–many pattern of influence. Suppose a whole squadron of bombers, with fighter escort, are on their way to destroy an extended target by carpet-bombing; a squadron of interceptors arrives to attack the bombers. In the ensuing dogfight, some of the interceptors are shot down. The remaining interceptors shoot down some of the bombers. The remaining bombers proceed to their assigned targets. Which parts of the target area get hit depends on which bombers get through. Thus various alterations of the dogfight would lead to various alterations of the destruction of the target area.

[23] Unless you alter Billy's throw so much that his rock arrives first, making Billy the preempting cause. In a context in which we're comparing Billy's throw and Suzy's, such alterations should be set aside as "too distant." I hope that the vagueness of the analysis at this point matches the vagueness of the analysandum, and so need not be regretted.

[24] Here I am indebted to Ned Hall.

[25] See my (1973b), pp. 4–19, on counterfactuals as variably strict conditionals.

[26] But we couldn't have had one of those alterations of $D$; because it would have to have been produced by a prior brain state that would have led the Neuroscientist to intervene.—True, but so what? We can still entertain them as counterfactual suppositions, and they can still constitute part of the pattern of influence from $D$ to $E$.

[27] In the case of the deadly double dose, an inclination to accept the counterexample may have a fourth source as well. Hall says that half of the double dose will cure Billy's nonfatal illness. So when we are told (truly, I take it) that Monday's dose causes Billy to survive, we're apt to hear a hint that it does so by curing his nonfatal illness. We too easily mistake the falsity of what's hinted for the falsity of what's actually said.

[28] Compare Lombard (1990), p. 197.

[29] Where an absence of spacetime itself is, there is nothing whatever there at all, relevant or otherwise. See my "Void and Object," chapter 10 in this volume, on voids as absences of spacetime, and on obstacles to the reification of absences.

30. *Pace* Bennett (1988) and Mellor (1995), pp. 156–162. Both Bennett and Mellor are willing to say that one fact causes another, where a fact either is or corresponds to a true proposition. Mellor does indeed deny that the two facts stand in a causal relation, where by "relation" he means a genuine universal existing in the world. (I deny that too.) But I am unappeased: even if causation is a "relation" only in some lightweight, unserious sense, still it shouldn't be said to relate propositions.

31. Helen Beebee, in "Causing and Nothingness," chapter 11 in this volume, states just this dilemma, but chooses the wrong horn of it.

32. The example comes from McDermott (1995a). It is further discussed in Collins, "Preemptive Prevention," chapter 4 in this volume. The suggestion that our wavering intuitions are governed by how far-fetched we find the possibility of the ball getting past the wall comes from Collins; but I have transplanted it from Collins's theory of causation as would-be dependence to my theory of causation as influence.

33. The puzzle was much discussed at Yale *circa* 1968; see Todes and Daniels (1975). It reappears in Sorensen (1999).