

# Preemptive Prevention<sup>\*</sup>

John Collins

Columbia University

As the ball flew towards us I leapt to my left to catch it. But it was you, reacting more rapidly than I, who caught the ball just in front of the point at which my hand was poised. Fortunate for us that you took the catch. The ball was headed on a course which, unimpeded, would have taken it through the glass window of a nearby building. Your catch prevented the window from being broken.

Or did it? Had you not made the catch, I would have caught the ball instead. My leaping to catch the ball made your catch redundant. Given my presence, the ball was never going to hit the window.

The example is a variant of one discussed by Michael McDermott. McDermott writes:

Suppose I reach out and catch a passing cricket ball. The next thing along in the ball's direction of motion was a solid brick wall. Beyond that was a window. Did my action prevent the ball hitting the window? (Did it cause the ball to *not* hit the window?) Nearly everyone's initial intuition is, 'No, because it wouldn't have hit the window irrespective of whether you had acted or not.' To this I say, 'If the wall had not been there, and I had not acted, the ball would have hit the window. So between us—me and the window—we prevented the ball from hitting the window. *Which*

---

<sup>\*</sup> Earlier versions of this paper were read to the Princeton Philosophy Colloquium on March 28<sup>th</sup> 1997 and to my Fall 1999 Metaphysics Seminar at Columbia. Thanks are due to Cian Dorr, Ned Hall, David Lewis, Cei Maslen, Laurie Paul, Jacob Rosen, and Jonathan Schaffer.

one of us prevented the ball from hitting the window—me or the wall (or both together)?’ And nearly everyone then retracts his initial intuition and says, ‘Well, it must have been your action that did it—the wall clearly contributed nothing.’<sup>1</sup>

These two cases, mine and McDermott’s, are certainly very similar. And what McDermott says about his example seems to me to be absolutely the right thing to say about mine. When the redundancy of your catch is raised, when it is pointed out that, given my presence, the window was never in any real danger of being broken, I reply (with McDermott) as follows. If neither of us had reached for the ball, then the ball would have hit the window. So between us—you and me—we prevented the ball from hitting the window. Which one of us prevented the ball from hitting the window—you or I (or both of us together)? Well, clearly it must have been you, for it was you and not I who took the catch. I contributed nothing.

Preemptive prevention is puzzling. My responses to these two cases differ, despite the fact that they share the same basic structure. I am happier to agree with what McDermott says about my version of the example than I am to concur with that same account of his own version of the story. When I stand between your catch and the window, I am happy to say that you prevented the window from breaking. It’s far less clear to me that that is the right thing to say when it is not just me but a solid brick wall that stands as the second line of defense. When it comes to McDermott’s version of the story, I find myself wanting to say the following things. (1) You prevented the ball from hitting the wall. (2) The wall would have prevented the ball from breaking the window. (3) Your catch prevented the wall from preventing the ball from breaking the window. (4) Likewise, the presence of the wall prevented your catch from preventing the ball from breaking the window. I am very reluctant to say, in this case, that your catch prevented the ball from breaking the window. Given that the wall was there, the window was never in any danger of being broken. The presence of the wall really does seem to make your catch irrelevant. At least, that’s how it seems to me.

---

<sup>1</sup> Michael McDermott, “Redundant Causation”, *British Journal for the Philosophy of Science* (1995) p.525.

Most, but not all, of the people I have asked shared the sense of ambivalence I experience when contemplating these two cases. Many, but not all, of them agreed with my judgements about the two examples. I think that these two cases of preemptive prevention pose a genuine puzzle. But perhaps those not convinced that the presence of the wall makes the catch irrelevant might be persuaded by an even more extreme variation on the same theme. Consider a third version of the preemptive prevention story. There is only one potential catchtaker this time, no brick wall, and no window. Suppose that at the moment you take the catch, the instantaneous velocity of the ball is a vector directed at a point one hundred million miles from the Earth at which Halley's comet will be located during its next but one swing through the solar system. Suppose that the magnitude of the instantaneous velocity is exactly right for a collision between the ball and the comet at that distant point at that future date (on the supposition that is, that the Earth and its gravitational field are absent). Did your catch prevent the ball from hitting Halley's comet? No.

## **Counterfactual Dependence**

Non-occurrences of events, as well as occurrences, may act as causes and may be produced as effects. Prevention is a matter of causing the non-occurrence of an event. To prevent the window from being broken is to cause the non-occurrence of a window-breaking.

There are two ways we might proceed here. We might extend our ontology to include events of non-occurrence, or we might continue simply to speak of the non-occurrence of events. The first approach leads to some difficulties in specifying when two events of non-occurrence are one and the same event. The second approach requires us to admit that causal relata may be other than events. Since I am dubious about the ontological status of events of non-occurrence, I propose to proceed as follows.

To any possible event  $e$  there corresponds the proposition  $E$  that the event  $e$  occurs. If  $e$  is any possible event, then the propositions  $E$  and  $\sim E$  will both be referred to as *propositions of occurrence*. Note that I am using the term in such a way as to include among the propositions of occurrence what might instead have been called propositions of non-occurrence; no problem, as long as you understand what I mean. Whenever  $C$  and  $E$  are propositions of occurrence,  $E$  would be true if  $C$  were true, and  $E$  would not have been true had  $C$  not been true, we say that  $E$  is *counterfactually dependent* on  $C$ . A sequence of true propositions of occurrence is said to be a *chain of counterfactual dependence* when each proposition in the sequence is counterfactually dependent on the proposition that precedes it. According to the *simple counterfactual analysis* of causation, a *causal chain* is just a chain of counterfactual dependence. One event (or absence) is a *cause* of another event (or absence) iff the proposition of occurrence for the former is linked to the proposition of occurrence for the latter by some causal chain.<sup>2</sup>

Suppose that  $C$ ,  $D$ , and  $E$  are true propositions of occurrence. We have *redundant causation* just in case: (i)  $E$  would not have been true had neither  $C$  nor  $D$  been true; (ii)  $E$  would still have been true if exactly one of the propositions  $C$  and  $D$  had been true and the other false. In the particular case where  $E$  is a proposition stating that a particular possible event did not occur, we have a case of redundant prevention.

Preemption is a special kind of redundancy. Following Lewis, let us mark the distinction in the following way. Suppose that events  $c$  and  $d$  are redundant causes or preventers of  $e$ . If it is clear, intuitively, that one of the two events  $c$  and  $d$  caused or prevented  $e$  while the other waited in reserve, then we shall speak of *preemption*. Those situation, on the other hand, in which  $c$  and  $d$  have an equal claim to having caused or prevented  $e$ , will be referred to as cases of *overdetermination*. The distinction is a vague one to be sure, yet there are clear cases of preemptive causation and prevention. The example with which the paper began is one such case.

---

<sup>2</sup> See David Lewis "Causation", in *Philosophical Papers* Vol. II (Oxford University Press, 1986) pp.159-172.

## Quasi-Dependence

The phenomenon of preemption stands in the way of a simple counterfactual analysis of causation. When it became clear that not all cases of preemption could be handled by the simple counterfactual analysis, Lewis introduced the notion of *quasi-dependence*. The idea was that whether or not a process is causal ought to depend only on the intrinsic nature of the process itself, and on the laws of nature. If a duplicate of some process is located in the same world, or in a world with the same laws of nature, call it an *isonomic duplicate*. Lewis's suggestion was that any isonomic duplicate of a causal process is a causal process. Call an isonomic duplicate of a counterfactually dependent chain a *quasi-dependent chain*. Then, according to the *quasi-dependence analysis* of causation, a *causal chain* is either a chain of counterfactual dependence or a quasi-dependent chain.<sup>3</sup>

But the quasi-dependence strategy doesn't work either. There are clear counterexamples to the claim that every isonomic duplicate of a causal process is a causal process, i.e. to the claim that causation is an intrinsic matter. One kind of counterexample involves cases of *causation by double prevention*: a cause produces an effect by preventing an event, which, had it not been prevented, would have prevented the effect.<sup>4</sup> But another kind of counterexample is provided by the phenomenon of preemptive prevention. Your catch prevents the window from breaking when it preempts my catch, but not when it merely prevents a collision with a brick wall from preventing the window breaking. Yet the only difference between these two cases has to do with features extrinsic to the simple process involving your catch. The process that includes the ball's flight, your catch and the window's not breaking is causal in the case where my hand was poised behind yours to take the catch, but it is *not* causal in the case where a brick wall is there instead of me.<sup>5</sup> We are assigning a different causal status to processes

---

<sup>3</sup> For details see David Lewis, "Postscripts to 'Causation'", in *Philosophical Papers*, Vol. II pp.193-212.

<sup>4</sup> For details of the examples, see Michael McDermott, "Redundant Causation", and Ned Hall, "Two Concepts of Causation", forthcoming.

<sup>5</sup> If you don't agree with me about this pair of examples, consider the Halley's comet case instead.

that are isonomic duplicates. So whether or not a process is causal is not a matter intrinsic to it.

Cases of preemptive prevention deserve to be added to the standard set of hard cases against which theories of causation are tested.

## Dependence Prevention

Your catch prevented the window from breaking and also prevented me from catching the ball when I leapt. But my leap did some preventing as well. It was my leap that prevented the window's not breaking from being counterfactually dependent on your catch. My leap was a *dependence preventer*. I think that the reason we say your catch prevented the window from breaking, despite the absence of counterfactual dependence, is that it was *only* my happening to leap when I did that prevented the not-breaking from depending on the catch. Sometimes, when  $e$  is not counterfactually dependent on  $c$ , we count  $c$  as a cause of  $e$  when there would be dependence were it not for some dependence preventer.

Of course this is not always so. We might recast the definition of redundancy as follows:  $c$  and  $d$  redundantly cause  $e$  iff (i)  $c$  prevents  $e$  from being counterfactually dependent on  $d$ ; and (ii)  $d$  prevents  $e$  from being counterfactually dependent on  $c$ . The redundant causes feature symmetrically in this definition. This symmetry creates a problem. For just as my leap prevented the not-breaking from depending on your catch, so did your catch prevent the not-breaking from depending on my leap. In cases of preemption not only does the preempted backup prevent the effect from depending on the preempting cause; the preempting cause also prevents the effect from depending on the preempted backup. Yet we judge the causal roles of preempter and preempted differently.

Might we account for this by distinguishing varieties of dependence prevention?

I once hoped so. For there are two ways of preventing counterfactual dependence and in many cases the distinction between these two kinds of dependence prevention matches the contours of our causal judgements. Consider some actual chain of counterfactually dependent events. The first way in which this chain of counterfactual dependence might have been prevented is if something had prevented one or more events in the chain from occurring. If you don't have a chain of actually occurring events, then you don't have a counterfactually dependent chain of actually occurring events. But an actual chain of events might have been prevented from being a chain of counterfactual dependence without any event in that chain itself being prevented. The second way of preventing dependence leaves the chain of events fixed and simply removes the counterfactual dependence. Call an event which prevents dependence in this second way a *pure dependence preventer*.

We may use this distinction to break the symmetry in our example of preemptive prevention. My leap and your catch were both dependence preventers, but only my leap was a *pure* dependence preventer. My action prevented the window's not breaking from being counterfactually dependent upon your reaching out to catch the ball, and it did so without preventing any occurrence (or non-occurrence) in the sequence initiated by your reaching out. Your catch also prevented the window's not breaking from being counterfactually dependent upon my leap, but it did so only by preventing the occurrence of a key event in the chain that would otherwise have followed from my action. In particular: your catch prevented me from catching the ball. This suggests that:

*c* is a cause of *e* iff there is a chain of counterfactual dependence linking *c* to *e*, or there would be such a chain were it not for some pure dependence preventer.

This needs adjustment. To simplify things in the preceding paragraphs I wrote as though all of our causal relata—all of the elements of our causal chains—had to be actually occurring events. But as we have already noted, non-occurrences can also cause

and be caused. A more accurate expression of the above idea would be couched in terms of propositions of occurrence. Thus:

A *causal chain* is a chain of true propositions of occurrence which is a chain of counterfactual dependence or *would be* were it not for the truth of some proposition of occurrence.

Then finally we might say that one event *c* (or the absence of the event *c*) is a cause of another event *e* (or the absence of *e*) iff there is a causal chain linking the proposition that *c* occurs (or doesn't occur) with the proposition that *e* occurs (or doesn't occur). Think of this as the *would-be dependence* analysis of causation.<sup>6</sup>

The would-be dependence analysis handles well many of the hard cases of preemption that plagued the quasi-dependence theory. It has another nice feature as well. It allows a solution to our puzzle about preemptive prevention. And the solution it allows is an attractive one, for it not only enables us to distinguish cases like mine from McDermott's; it also explains our feeling of ambivalence about such examples.

I say that the difference between McDermott's example and mine is this: the counterfactual assumption of the absence of the pure dependence preventer in his story is *more far-fetched* than the corresponding assumption of absence in mine. It does not require much of a stretch to suppose that I simply get my timing slightly wrong, so that when I leap, I do so at not quite the right moment to be ready to take the catch. It is more far-fetched, on the other hand, to suppose that the brick wall be absent, or that the ball would miraculously pass straight through it.

So let us adjust the would-be dependence analysis accordingly:

---

<sup>6</sup> The material in this section was developed independently in late 1996 and presented in an earlier version of this paper read to the Princeton Philosophy Colloquium in March 1997. In discussion following that paper Cian Dorr drew my attention to a version of the "would-be dependence" analysis due to Jonardon Ganeri, Paul Noordhof, and Murali Ramachandran. See their papers "Counterfactuals and Preemptive Causation", *Analysis* 56.4 (October 1996) pp.219-225 and "For a (Revised) PCA-Analysis", *Analysis* 58.1 (January 1998) pp.45-47.

A *causal chain* is a chain of true propositions of occurrence which is a chain of counterfactual dependence or *would be* were some true proposition of occurrence false *in some not too far-fetched way*.

This analysis is vague, since it is a vague matter which alternatives to actuality should count as being “too far-fetched” to be taken into account. But this vagueness is a strength of the analysis, for it properly matches the vagueness of the analysandum. Ambivalent intuitions about examples like McDermott’s stem from uncertainty about which suppositions count as being too far-fetched. In the right frame of mind I might be persuaded to suppose away a solid brick wall. Some competent speakers are happy to do so, McDermott among them. But fetch mere possibility from far enough, and none will acquiesce in supposing the dependence preventer absent. The Earth, its atmosphere, and its gravitational field prevent the ball’s non-collision with Halley’s comet from depending on your catch. Far fetched indeed the supposition that you catch the ball in the Earth’s absence.

## **Trumping and Influence**

But not every chain that would be dependent were it not for some pure dependence preventer is a causal chain. That is one of the lessons to be learned from Jonathan Schaffer’s discovery of *trumping preemption*.<sup>7</sup> So trumping preemption is bad news not only for the quasi-dependence theory, but for the would-be dependence account as well. I now think of the latter as the *would-have-been* theory; I no longer defend it.

In one of Schaffer’s examples the laws of magic are such that what happens at midnight is determined by the first spell cast the previous day. At noon Merlin casts the first spell of the day: a spell to turn the prince into a frog. At six that evening Morgana casts the second (and last) spell of the day: a spell to turn the prince into a frog. At

---

<sup>7</sup> Jonathan Schaffer, “Trumping Preemption”, forthcoming.

midnight the prince turns into a frog. The transfiguration is caused by Merlin and not Morgana, yet it would be dependent on Morgana's spell were it not for Merlin's trumping preemption, and Merlin's preempting spell is a *pure* dependence preventer if we allow that magical spells act without causal intermediaries.

David Lewis has responded to the examples of trumping preemption by giving up on the idea of providing a counterfactual analysis of *c is a cause of e*. His hope now is to provide in its place a counterfactual theory of *causal influence*. The idea is, roughly, this. Our original notion of the counterfactual dependence of one event on another was a notion of *whether-whether* dependence; *e* is dependent on *c* in this sense just in case whether or not *e* occurs depends on whether or not *c* occurs. But there are further varieties of dependence. Lewis's suggestion is that we think of degree of causal influence as being a matter of the degree to which *whether, when, and how* one thing happens depends upon whether, when, and how something else happens.<sup>8</sup> Thus the asymmetry in the Merlin-Morgana example is revealed by the fact that Merlin, but not Morgana, has a causal influence on what happens at midnight. Had Merlin cast a different spell at noon, something different would have happened at midnight, but what happened at midnight was in no way dependent upon whether, when, or how Morgana acted later.

This new theory of causal influence amounts to a change of topic. Lewis is not offering a new answer to the old question: "what is it for this event to be a cause of that event?" He is rather, offering an answer to a quite new question: "what is it for this event to have a causal influence on that event?" Certainly one event can have a causal influence on another without being among its causes. An example is provided by the story of the *Poison and the Pudding*.<sup>9</sup> If a poison kills its victim more slowly and painfully when taken on a full stomach, then the victim's eating pudding before he drinks the poisoned potion has a causal influence on his death, since the time and manner of the

---

<sup>8</sup> For the full details of this proposal, see David Lewis, "Causation as Influence", forthcoming.

<sup>9</sup> David Lewis, "Postscripts to 'Causation'", in *Philosophical Papers*, Vol. II pp.198-99.

death depend counterfactually on the eating of the pudding. Yet the eating of the pudding is not a cause of his death.

The relation of cause and effect does not reduce in any obvious way to degree of causal influence. Dependence of the *whether-whether* kind appears to play a special role in our judgements of what is a cause of what. As it stands, the new theory lacks the resources to identify an event's causes from among those things that merely had some causal influence on it.

Furthermore, as Jacob Rosen has pointed out to me, there are cases of trumping preemption that appear to be counterexamples to the necessity of Lewis's account of causal influence. Changing Schaffer's story slightly, let us suppose that the laws of magic distinguish the powers of a wizard from those of a witch. As a wizard, Merlin's options are very limited indeed. There is only one spell he may cast (standard prince-to-frog) and only one time of day he may cast it (noon). No such limitations apply to Morgana. Let the rest of the story remain the same. The prince's transfiguration is now in no way dependent on whether, or when, or how Merlin acts. So according to Lewis's account, Merlin's spell has no causal influence on what happens at midnight. Yet I would claim, with Rosen, that it is still Merlin's spell, and not Morgana's, that causes the prince to turn into a frog.<sup>10</sup>

## Masking and Finkish Trumping

Time now to turn our attention to another failed counterfactual analysis. To say that a glass window is fragile, to say that is, that the window is disposed to shatter when struck, would seem to be equivalent to saying that *if the window were struck, it would shatter*. However, as C.B.Martin noticed long ago, this simple counterfactual analysis of dispositions is false. For dispositions, like other properties, may be gained or lost, and

---

<sup>10</sup> Jacob Rosen, personal communication, November 1999.

the conditions under which a disposition would be lost might be exactly the disposition's conditions of manifestation. Call such dispositions *finkish*.<sup>11</sup>

Further problems for the simple counterfactual analysis are presented by dispositions that *mask* other dispositions.<sup>12</sup> To say that I mean *addition* when I use the word 'plus', is to ascribe certain dispositions to me; for example, the disposition to give the sum of two numbers  $x$  and  $y$  when asked "What is  $x$  plus  $y$ ?" But of course it is not true for all values of  $x$  and  $y$  that were I asked "What is  $x$  plus  $y$ ?" I would respond by giving their sum. That is because the disposition that constitutes my meaning addition by 'plus' is masked by various other dispositions of mine: my disposition to make mistakes, my disposition to decline to answer annoying questions, my disposition to get old and die ...

I say: a masked disposition may fail to manifest itself even though it remains present and its conditions of manifestation are satisfied. Some disagree. For example David Lewis writes:

the first problem we face in analysing any particular dispositional concept ... is the problem of specifying the stimulus and the response correctly.

We might offhand define a poison as a substance that is disposed to cause death if ingested. But that is rough: the specifications both of the response and the stimulus stand in need of various corrections. To take just one of the latter corrections: we should really say 'if ingested without its antidote'.<sup>13</sup>

---

<sup>11</sup> C.B.Martin, "Dispositions and Conditionals", *The Philosophical Quarterly* 44 (1994) pp.1-8. It should be noted that Martin first came up with the idea of finkish dispositions in 1957.

<sup>12</sup> For masking, and the related phenomena of altering and mimicking, see Mark Johnston, "Objectivity Refigured: Pragmatism Without Verificationism", Appendix 2: "Complexities in the Notion of a Disposition", in *Reality, Representation and Projection*, Haldane and Wright (eds.) Oxford University Press (1993)

<sup>13</sup> David Lewis, "Finkish Dispositions", in *Papers in Metaphysics and Epistemology*, Cambridge University Press (1999).

Lewis may well be right about this particular case; it is probably false to say that the disposition of a poison to kill is masked by its antidote. But I deny the general claim that Lewis's example is intended to support. For to accept that claim would be to rob ourselves of one of the most crucial conceptual resources provided by the notion of a disposition. Some of our most important dispositional concepts are concepts of dispositions that are components of other dispositions. These component dispositions may have conditions of manifestation that fail to correspond to the antecedent of any appropriate counterfactual conditional. That is how, for example, we make sense of the dispositions of belief, desire, and meaning as component dispositions of an agent's total behavioral disposition.<sup>14</sup> And that is how we may reply to a Kripkensteinian skepticism about the notion of meaning.<sup>15</sup>

David Lewis has offered a reformed counterfactual analysis of dispositions.<sup>16</sup> The reformed theory relies on the thesis that a disposition must be absent whenever the conditions for manifestation are satisfied and yet the manifestation fails to occur. But if masking works the way I claim it does, that thesis is false.

Prior, Pargetter, and Jackson argue that every disposition must have a causal basis.<sup>17</sup> Lewis agrees with them. A *causal basis* for the window's disposition to shatter when struck is said to be some intrinsic property of the window that would cause it to shatter were it struck. But if this counterfactual analysis of the notion of a causal basis is correct, then examples of *finkish trumping* establish that not every disposition has a causal basis. Consider an example that involves preemption or *mimicking* rather than masking. Suppose that the window is fragile, but that a playful sorcerer watches and waits, determined that if ever the window is struck he will, quick as a wink, cast a spell to shatter the window. Suppose that the spell would work not by changing any of the

---

<sup>14</sup> See my "Indeterminacy and Intention", in *The Philosophy of Donald Davidson*, The Library of Living Philosophers Vol. XXVII, Lewis Hahn (ed.) Open Court (1999) pp.501-528.

<sup>15</sup> See Saul Kripke *Wittgenstein on Rules and Private Language*, Harvard University Press (1982), to which the 'plus' example above refers. Kripke's rejection of a dispositional solution to the skeptical paradox presupposes the counterfactual analysis of dispositions.

<sup>16</sup> David Lewis, "Finkish Dispositions".

<sup>17</sup> Elizabeth Prior, Robert Pargetter and Frank Jackson, "Three Theses about Dispositions", *American Philosophical Quarterly*, 19 (1982) pp.251-253.

intrinsic properties of the window, nor by cutting any causal chain initiated by the striking and any intrinsic property of the window, but simply by the trumping preemption of whatever ordinary causal process would lead an unenchanted duplicate of the window to shatter when struck. Then the window would shatter were it struck, but its shattering would not be caused by any intrinsic property. So, if the counterfactual analysis of the notion of a causal basis is correct, the window has the disposition to shatter when struck, but this disposition has no causal basis.

According to a (partially) reformed counterfactual analysis of what it is for the window to be fragile:

The window is disposed to shatter when struck iff, the window has an intrinsic property *B* such that, if the window were struck and yet retained *B*, then the striking and the window's having property *B* would jointly cause the window to shatter.<sup>18</sup>

If this correct, then the window in the finkish trumping example is *not* disposed to shatter when struck, despite the fact that it is an isonomic duplicate of a window that is so disposed. Defenders of the reformed counterfactual analysis must deny that dispositions are an intrinsic matter.

---

<sup>18</sup> For the full reformed analysis, see David Lewis "Finkish Dispositions", pp.148-149. I am simplifying matters here by omitting two features of Lewis's final proposal, but those features are not relevant to the point I am making here