ON LEWIS'S COUNTERFACTUAL ANALYSIS OF CAUSATION

Wylie Breckenridge

For some time, David Lewis has been trying to find a satisfactory counterfactual analysis of causation. In this essay I will discuss four of his most significant attempts, from the first, offered in 1973, to his most recent, offered in 1999. My approach is as follows. For each analysis I will (i) describe the motivating idea behind it, (ii) state it, (iii) describe why Lewis thinks it is an improvement on its predecessors (except for the first), (iv) present the main objections that have been levied against it, (v) consider ways of defending it against those objections, both Lewis's and others, and finally (vi) summarise Lewis's position on the objections and on the degree to which the analysis succeeds. The presentation is intended to be impartial, and I will draw no conclusions other than Lewis's own about the success or failure of the analyses or of his project as a whole. Nevertheless, I'd be happy for the repeated failures that this essay describes to be seen as premises for an abductive inference to the conclusion that I lean towards – that a satisfactory counterfactual analysis of causation cannot be found.

1. Lewis's Aim

It is important to be as clear as possible about what Lewis is trying to achieve. His aim is to give a reductive conceptual analysis of causation in terms of counterfactuals. That is, (i) he wants to explain the *meaning* of the concept we express in causal statements like "The derailment was caused by my placing a brick on the track", "The April rain caused there to be an unburnt forest in June", and "Smoking causes lung cancer", (ii) he wants the explanation to be given in terms of counterfactuals – statements of the form "If so-and-so were the case then such-and-such would be the case", and (iii) he wants the explananda (i.e. the counterfactuals) to have meanings that can themselves be explained without appeal to causation. An analysis of the kind he is after might go something like this: "The derailment was caused by my placing a brick on the track" means that if my placing the brick on the track had not occurred then the derailment would not have occurred. As we shall see this simple analysis cannot be right, but it gives the flavour of what Lewis wants.

Although this is Lewis's ultimate aim, his more immediate aim has been less ambitious in two ways. First, although it seems that each of our causal statements expresses a relation, they don't all seem to relate the same kinds of things. Consider the three statements above. The first expresses a relation between events: the derailment, and my placing a brick on the track. The second expresses a relation between an event and a fact: the April rain, and there being an unburnt forest in June. The third expresses a relation between general kinds: smoking, and lung cancer. If different kinds of relata imply different relations then it would seem that we have at least three concepts of causation – call them *event* causation, *fact* causation and *general* causation. If that is right, then to give a complete analysis of causation Lewis has to give an analysis of each. It may be, however, that we have just *one* concept of causation, even though the words that we use suggest otherwise. It may be, for example, that we conceive of causation only as a relation between events, and that talk of causation between facts and general kinds is really just about that. Or it may be that we do have at least these three concepts but that one of them is more basic, so that, for example, our concepts of fact

causation and general causation can be explained in terms of our concept of event causation. These are interesting questions, but will not concern us here. Lewis's aim is analyse our concept of *event* causation - the concept that we express in statements that, like the first, relate two particular events. This may or may not be the same concept that we express in statements that relate facts or general kinds, but it is *this* concept whose meaning he wants to analyse.

Second, it is a necessary condition on "so-and-so" being an adequate *conceptual* analysis of "such-and-such" that it be an adequate *extensional* analysis – that "so-and-so" be true if and only if "such-and-such" is true. It is for this reason, for instance, that we reject "This object is a red piece of fruit" as giving the meaning of "This object is an apple": when said of a green apple the second is true but the first is false. Moreover, it is a necessary condition that this extensional adequacy extend to *all possible worlds*. We rule out "This animal has a heart" as being the meaning of "This animal has a kidney" not because one is true while the other is false in *this* world, but because one is true while the other is false in some possible worlds (in any world, in fact, where at least one animal has a heart but not a kidney, or vice-versa). So it is a necessary condition on an adequate conceptual analysis of causation in terms of counterfactuals that whatever set of counterfactuals Lewis proposes as the meaning of a given statement of event causation, they must be true if and only if the causal statement is true, not just in this world but in all possible worlds. It turns out that satisfying this condition has been difficult enough, and that has been Lewis's more immediate aim. Note that being an adequate extensional analysis is *not*, however, *sufficient* for being an adequate conceptual analysis, even if it is extensionally adequate in all possible worlds. "This is a triangle" and "This is a trilateral" are true of exactly the same shapes in all possible worlds, but it seems that neither gives the meaning of the other. So even if Lewis can achieve an adequate extensional analysis, it is not guaranteed to be an adequate conceptual analysis. That is a further and interesting question, but not one that will concern us here.

I will take Lewis's aim to be narrower in two more ways. Sometimes we think that one event causes another without *determining* it: I might assert, for example, that "My installing the radioactive source caused the bomb to explode", and yet believe there was some chance that the decay and hence the explosion would not have occurred. Sometimes we think that an *absence* of an event causes another: I might assert that "Not drinking any water was the cause of my getting a headache." Lewis has been aiming for his analysis to extend to both of these types of statement: ones in which the cause is thought to not determine the effect, and ones in which the cause (and/or effect) is thought to be the absence of an event. To simplify my exposition I will ignore that part of his aim. That will still leave us plenty to do.

From here on, then, I will take Lewis's aim to be that of giving an extensionally adequate, reductive counterfactual analysis of our concept of causation between events (ones that actually occur, not ones that fail to occur) in a deterministic world – the concept expressed in causal statements like "The derailment was caused by my placing a brick on the track", but not necessarily (although possibly) in statements like "The April rain caused there to be an unburnt forest in June", "Smoking causes lung cancer", "My installing the radioactive source caused the bomb to explode", and "Not drinking any water was the cause of my getting a headache."

At this point I would like to consider and then put aside one issue arising from Lewis's aim. It might be objected that he must fail in his aim of giving a *reductive* analysis of causation in terms of counterfactuals, because our concept of the former is more basic than our concept of the latter so if there is to be any reductive analysis it can only be of counterfactuals in terms of causation. That is, it might be objected that there can be no analysis of counterfactuals that is itself causation-free. But Lewis has offered an analysis that takes some of the force out of this objection.[1] First he says that "If A were true then C would be true" is true if and only if there is a possible world in which A and C are true that is more similar to the actual world than any world in which A is true and C is false. Next he says that world W_1 is more similar to the actual world than world W_2 if and only if

(1) it has fewer big, widespread, diverse violations of the laws of the actual world than W_2 does;

OR

(2) it has equally big, widespread, diverse violations of the laws of the actual world as W_2 does, but its spatio-temporal region of perfect match with matters of particular fact in the actual world is greater than that of W_2;

OR

(3) it has equally big, widespread, diverse violations of the laws of the actual world as W_2 does, it has an equally extensive spatio-temporal region of perfect match with matters of particular fact in the actual world as W_2 does, but it has fewer small, localised, simple violations of the laws of the actual world than W_2 does.

Finally, he says that a law is a necessary generalisation.

Whether or not this is an adequate analysis of counterfactuals and whether or not it really is causation free will not concern us here. The important point for our purposes is that Lewis's aim of providing a *reductive* analysis of causation in terms of counterfactuals cannot be immediately dismissed as mistaken. We will not concern ourselves with that part of Lewis's aim from here on, but only with his aim of providing an *extensionally adequate* analysis.

2. Methodology

Each of Lewis's analyses makes a claim of the form "C is a cause of E if and only if P", where P is some proposition about non-causal facts and counterfacts. Its extensional adequacy is tested by looking for counterexamples - courses of events in which "C is a cause of E" is true but "P" is false, or in which "C is a cause of E" is false but "P" is true. Cases like the first show that the analysis suffers from *undergeneration* – it rules out some cases of causation. Cases like the second show that it suffers from *overgeneration* – it rules in some cases of non-causation. Both are counterexamples to the analysis and show that it is inadequate.

We are engaged in conceptual analysis, so testing the truth of "C is a cause of E" and "P" is a matter of asking ourselves (and each other) whether or not we *think* they are true, are prepared to assert them, find them intuitive, and so on. If we think that one is

[1] Lewis 1973, 1979.

true but not the other then we have an *apparent* counterexample to the analysis – apparent, because there is more to these being true than someone thinking that they are true, and it does not necessarily follow that the analysis is at fault. "C is a cause of E" is true if and only if it is (or would be) assented to by anyone who (i) is applying the target concept of causation, and (ii) is applying that concept correctly. Similarly, "P" is true if and only if it is (would be) assented to by anyone who (i) is applying the concept of the counterfactual that the analysis intends to be applied, and (ii) is applying that concept correctly. When faced with an apparent counterexample – a case in which someone is prepared to assent to "C is a cause of E" but not "P", or vice-versa – the analysis can be defended by arguing any of these four things: (i) the person is applying a concept of causation different to the target one, (ii) she is applying the target concept but is doing so incorrectly, (iii) she is applying a concept of the counterfactual different to the one intended by the analysis, or (iv) she is applying the intended concept of the counterfactual but is doing so incorrectly. In what follows, we shall see each of these lines of defence being used to defend Lewis's analyses. We shall see others as well, some of which seem to be good things to say, some of which do not, but all of which show that testing analyses is not a cut-and-dried matter.

One final point. Lewis's target concept is that of *event* causation – the one we express in statements that relate events, but may not express in statements that relate facts or general kinds. Of a given course of events we may assent to or dissent from statements of all three kinds, but only those that are about events are relevant to assessing to the adequacy of Lewis's analyses. The importance of keeping this in mind will become clear later.

3. The First Analysis: Counterfactual Dependence

Lewis's first analysis is inspired by David Hume's second definition of causation: "we may define a cause to be an object followed by another, … where, if the first had not been, the second never had existed."[2] Lewis puts it this way. Say that event E *counterfactually depends* on event C if and only if if C had not occurred then E would not have occurred. Then:

> L1: Event C is a cause of event E if and only if E counterfactually depends on C.

We need not look to complicated test cases to find trouble for L1. Here is an apparent counterexample:

> The Word. Tom writes a word.

Why is this an apparent counterexample? We are prepared to assent to "if Tom's writing the word had not occurred then Tom's writing the word would not have occurred", but not to "Tom's writing the word was a cause of Tom's writing the word", and we are prepared to assent to "if Tom's writing the first letter of the word had not occurred then Tom's writing the word would not have occurred", but not to "Tom's writing the first letter of the word was a cause of Tom's writing the word." This is contrary to L1, which

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

claims that if we are prepared to assent to the first statement in each case then we will be prepared to assent to the second. How might L1 be defended against this? It is difficult to see how - our concept of causation seems to be such that an event cannot be caused by itself or any part of itself, but our concept of the counterfactual seems to be such that an event *can* counterfactually depend on itself and part of itself. Lewis accepts this is as genuine problem for L1 – that it suffers from overgeneration if we allow C and E to be the same event, or if we allow C to be a part of E. In response, he restricts the class of event pairs to which it is intended to apply:

L1': Event C is a cause of event E if and only if C and E are distinct events and E counterfactually depends on C.

Lewis now takes it as read in all of his analyses that C and E are required be distinct events and does not explicitly say so. We will do the same.

Another apparent counterexample:

The Thermometer. The air temperature rises and the mercury in the thermometer expands.

Why is this an apparent counterexample? We are prepared to assent to "if the expansion of the mercury had not occurred then the rising of the temperature must not have occurred", but not to "the expansion of the mercury was a cause of the rising of the temperature." Is this a genuine counterexample? Lewis thinks not. He defends L1 by arguing that anyone who assents to the first is using a different concept of the counterfactual conditional than the one he intends to be used. The counterfactual expressed here is what he calls a *backtracking* counterfactual. Its distinctive feature is the word "must" – without it, or some semantically equivalent expression like "would have had to have", we would no longer be prepared to assert the counterfactual: we would not be prepared to assert, for example, that "if the expansion of the mercury had not occurred then the rising of the temperature would not have occurred" – intuitively it sounds wrong. The use of such backtracking counterfactuals is already ruled out by the precise wording of L1, but we could add an extra clause to be sure:

L1'': Event C is a cause of event E if and only if E counterfactually depends on C (with no backtracking allowed).

Lewis also now takes it as read in all of his analyses that backtracking counterfactuals are not allowed and does not explicitly say so. Again, we will do the same.

Another apparent counterexample:

Socrates's Death. Socrates dies and Xanthippe (his wife) becomes a widow.

How is this argued to be a counterexample? Some people are inclined to think that if Socrates's death had not occurred then Xanthippe's becoming a widow would not have occurred, but that Socrates's death was *not* a cause of Xanthippe's becoming a widow –

to them, Xanthippe's becoming a widow is not the kind of event that gets *caused*. Lewis does not seem troubled by this, noting in passing that L1 needs the right kinds of events. I think he could say that such people are violating the antecedent conditions of L1. Their problem is with the phrase "Xanthippe's becoming a widow." It either refers to an event or it does not. If it does not, then the example is excluded from analysis by L1 because it requires E to be an event. If it does, then arguably it can only refer to the very same event that "Socrates's death" does (i.e. Xanthippe's becoming a widow and Socrates's death are identical events). Why? Suppose we say that it does refer to an event. It seems reasonable to think that our concept of an event is such that (E1) every event occurs in some region of space and in some interval of time, and that (E2) no two distinct events can occupy the same region of space during the same interval of time (i.e. that events can be identified by the space-time region that they occupy) (at least, it seems reasonable to think that E1 and E2 are true of events like Socrates's death and Xanthippe's becoming a widow, even if not of *all* events). Where and when does Xanthippe's becoming a widow occur? There is one natural choice for *when* it occurs – the time of Socrates's death. But there are two natural choices for *where* it occurs – the place where Socrates dies, or the place where Xanthippe is at that time. Suppose we say the latter. Suppose also that at the same time Xanthippe's sister gives birth, so that Xanthippe becomes an aunty. To be consistent we should say that Xanthippe's becoming an aunty occurs at the same time and place as Xanthippe's becoming a widow. Then according to E2 they must be identical events: Xanthippe's becoming an aunty is the same as Xanthippe's becoming a widow. This is an unattractive position, forced upon us if we say that Xanthippe's becoming a widow occurs where *she* is rather than where *Socrates* is. So we ought to say the latter instead. But then we have Xanthippe's becoming a widow occurring at the same time and in the same place as Socrates's death, so, according to L2, they must be identical events. But then the example is again excluded from analysis by L1 because it requires C and E to be distinct events (remember: this is now an unstated condition). So whether or not "Xanthippe's becoming a widow" refers to an event, the example does not meet the antecedent conditions of L1 and cannot be a counterexample.

Another apparent counterexample:

> The Push. I push Jones in front of a truck, which hits him and kills him. If I had not done so, he would have been hit and killed by a bus.

How is this argued to be a counterexample? We are inclined to think that my pushing Jones was a cause of Jones's death, because my pushing Jones was a cause of the bus hitting Jones and the bus hitting Jones was, in turn, a cause of Jones's death. We seem to think, at least in this case, that causation is *transitive* – if C is a cause of D and D is a cause of E then C is a cause of E. We are not, however, inclined to think that if my pushing Jones had not occurred then Jones's death would not have occurred, because Jones would have died anyway. This despite the fact that we are inclined to think that if my pushing Jones had not occurred then the bus hitting Jones would not have occurred, and if the bus hitting Jones had not occurred then Jones's death would not have occurred (remember: no backtracking allowed, so it cannot be objected that if the bus hitting Jones had not occurred then Jones's death would still have occurred because my pushing Jones must not have occurred). Counterfactuals are not invariably transitive (as this example shows).

Lewis accepts this as a genuine problem for L1, and responds by offering the analysis that we will look at next. But before we do, we should consider whether or not Lewis is right to think that this is a counterexample. How might L1 be defended against it? There are at least two ways. The first is to say that anyone who assents to "if my pushing Jones had not occurred then Jones's death would still have occurred" should not also assent to "my pushing Jones was a cause of his death" – that this would be a misapplication of the target concept of causation. McDermott has a good reply:

> If the bus had not been there, and I had not pushed, Jones would not have died. So between us – me and the bus – we caused his death. Which one of us caused his death – me or the bus (or both together)? … It must have been [my] push that did it – the bus clearly contributed nothing.[3]

The second is to say that it is simply wrong to think that if my pushing Jones had not occurred then Jones's death would still have occurred, because his death at the hands of the truck would have been a *different* death. This is to say that anyone who thinks the former is misapplying the concept of the counterfactual conditional. This defence seems to be stronger. One response might be to point out that had I not pushed Jones then Jones would still have died. The truth of that statement is hard to deny, and it seems to be saying the same thing. But it is importantly different – it expresses a counterfactual relation between two *facts*, that I pushed Jones and that Jones died, not between the occurrence of two *events*, my pushing Jones and Jones's death. L1 is a claim about counterfactual relations between the occurrence of events, and these are the only counterfactuals that are relevant to its assessment. Although it is clear that if I had not pushed Jones then Jones would still have died, it is certainly not so clear that if my pushing of Jones had not occurred then Jones's death (the event that actually occurred) would still have occurred (and not been replaced by a *different* event that would also have made it true that Jones died). So this response is not acceptable and the defence stands.

Here is a case that might weaken this second defence:

> The Bell. Tom fires a bullet which hits bell A which rings and makes Harry jump with fright. If Tom had not fired the bullet, Dick would have rung bell B which would have made Harry jump with fright instead.

As for The Push, we are happy to say that Tom's firing of the bullet caused Harry's jumping with fright, but not that if the former had not occurred then the latter would not have occurred - if Tom's firing had not occurred then Harry's jumping with fright would *still* have occurred. Moreover, by adjusting the example as necessary we can make Harry's jumping with fright at the hands of bell B as similar as we like to his jumping with fright at the hands of bell A. We can, for instance, make it so that only the sound of the bells makes Harry jump and we can arrange them so that no matter which one rings the same pattern of sound waves arrives at Harry's ears at the same time (for *any* standard of sameness that is required). Then what could it matter to when, where and how he jumps which bell rings? That should be enough to weaken the second defence to the previous example.

[3] McDermott 1995b, pp 524-5.

A stubborn defender of L1 could refuse to accept, of course, that we can make Harry's jumping with fright *exactly* the same in both cases, and claim that therefore his actual jumping with fright is indeed counterfactually dependent upon Tom's firing. But this argument is invalid. Even if we think the two jumps are slightly different it doesn't follow that we will think they are different events. It doesn't follow because there are many cases in which we think the very same event could have occurred in quite different ways, and in quite different places and at quite different times. For the argument to be valid it needs an extra premise that claims that this particular event (the jumping with fright that actually occurred) could not have occurred other than exactly as it did. But then the stubborn defender owes us a reason for thinking that this extra premise is true, when it is common-sensibly false for so many other events. So it seems that both attempted defences of L1 fail, and that Lewis is right to that it is an inadequate analysis.

In summary, we have looked at five apparent counterexamples to L1. The Word is accepted by Lewis as genuine and prompted him to add the (now unstated) condition that C and E be distinct events. The Thermometer is rejected by Lewis, but prompted the clarification that backtracking counterfactuals are not allowed. Socrates's Death is accepted by Lewis as showing that L1 needs C and E to be the right sorts of events, but I have suggested that requiring them to be distinct is already sufficient. Finally, The Push and The Bell are both accepted by Lewis as showing that L1 fails in some cases where causation is transitive but counterfactual dependence is not, and have prompted the analysis we will look at next.

4. The Second Analysis: Chains of Counterfactual Dependence

We have seen that cases like The Push and The Bell are a problem for L1 because transitivity holds for causation but not for counterfactual dependence. In response, Lewis just builds transitivity in:

> L2: C is a cause of E if and only if there is a chain of counterfactual dependence from C to E. (That is, for some n >= 1 there is a sequence of distinct events (C = D_1, …, D_n) such that D_2 counterfactually depends on D_1, D_3 counterfactually depends on D_2, …, and E counterfactually depends on D_n).

How is L2 meant to avoid the problem faced by The Push and The Bell? In the case of The Push, although we might *not* be inclined to think that Jones's death was counterfactually dependent on my push, we *are* inclined to think that there is a chain of counterfactual dependence between them: (my pushing Jones, Jones's collision with the bus, Jones's death) will do – if my pushing Jones had not occurred then Jones's collision with the bus would not have occurred, and if Jones's collision with the bus had not occurred then Jones's death would not have occurred (remember: no backtracking). In the case of The Bell, although we might *not* be inclined to think that Harry's jumping with fright was counterfactually dependent on Tom's firing of the bullet, we *are* inclined to think that there is a chain of counterfactual dependence between them: (Tom's firing the bullet, the ringing of bell A, Harry's jumping with fright) will do. So L2 seems to work in both cases. Note that by moving from L1 to L2, Lewis has not

created any new problems at the hand of the other cases we have considered: The Word and Socrates's Death are handled by the continued condition that C and E be distinct events, and The Thermometer is still handled by backtracking counterfactuals being not allowed. So now we have an analysis that is countered by none of our current test cases.

By allowing *chains* of counterfactual dependence, L2 is committed to causation being invariably transitive – to saying that in every case in which we think that C is a cause of D and that D is a cause of E we will think that C is a cause of E. Why? If we think that C is a cause of D then, according to L2, we think that there is a chain of counterfactual dependence from C to D, and if we think that D is a cause of E then we think that there is a chain of counterfactual dependence form D to E, so we must think that there is a chain of counterfactual dependence from C and E (the concatenation of the above two chains, at least) and hence that C is a cause of E.

It has been objected that this is a problem for L2, because we do *not* think that causation is invariably transitive. Lewis presents a list of so-called *counterexamples to transitivity*, of which I will consider just three.

> Forest Fire. In April there is rain. In May there is lightning, but no bushfire because the forest is wet from the April rain. In June there is more lightning, and because the forest has dried off there is a bushfire.

Lewis says that the April rain caused there to be an unburnt forest in June, which in turn caused the June fire, so if causation is invariably transitive then we must conclude (counterintuitively) that the rain caused the fire.

> Dog-Bite. My dog bites off my right forefinger. Next day I have occasion to detonate a bomb. I do it the only way I can, by pressing the button with my left forefinger. If the dog-bite had not occurred, I would have pressed the button with my right forefinger. The bomb duly explodes.

Lewis says that the dog-bite caused me to press the button with my left forefinger which in turn caused the explosion, so if causation is invariably transitive then (counterintuitively) the dog-bite was a cause of the explosion.

> Shock C. A and B each has a switch with two positions, Left and Right. To start, both are in Left position. A has first turn. He can either move his switch to Right or do nothing. B then has a turn. He can either move his switch to Right or do nothing. The power is then turned on. If both switches are in Left position, or both in Right position, C gets an electric shock. On this occasion, A, seeing that B's switch is in Left position and wanting to save C, moves his switch to Right. B wants C to get a shock so he responds by moving his switch to Right also. C duly gets a shock.

Lewis says that A's moving his switch Right caused B to move his switch Right which in turn caused C to be shocked. If causation is invariably transitive then (counterintuitively) A's failed attempt to prevent the shock was among its causes.

Lewis's response to these is that in each case transitivity *does* succeed, that to think otherwise is to misapply the target concept of causation. This is what he says (by

"Black's move" he means the first event in each case – the April rain, the dog-bite and A's move - and by "Red's victory" he means the last one – the forest fire, the explosion and the shock):

> 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 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.[4]

It sounds to me like Lewis is saying that Black's move causes something to be a cause of Red's victory and therefore is itself a cause of Red's victory, which is dangerously close to circularity: it sounds like he is defending the transitivity of causation by appealing to the transitivity of causation. Lewis admits to feeling some ambivalence about this response, but believes that it can be explained away (for details see Lewis 1999, pp. 31-32). Rather than considering how, I will offer another way of defending the transitivity of causation against these proposed counterexamples.

It is important to keep in mind two things. First, the issue here is not the extensional adequacy of L2. It is whether or not these are cases in which we intuitively think that transitivity of causation fails, independently of any proposed analysis. So to assess them we can only appeal to what we are intuitively inclined to say, and not to what L2 or any other analysis might have us say. Second, L2 commits Lewis to the transitivity of *event* causation, so that any genuine counterexample to this must be a situation in which we think that *event* C is a cause of *event* D, that *event* D is a cause of *event* E, but that *event* C is not a cause of *event* D. If we happen to find a counterexample to the transitivity of fact or general causation then we cannot assume that we have thereby found a counterexample to the transitivity of event causation. The importance of keeping these in mind will become clear in what follows.

First, the case of Forest Fire. What does Lewis say the April rain causes? "There to be an unburnt forest in June". But that refers to a *fact*, not an *event*, and we are looking for counterexamples that involve event causation. What *event* should we put in place of this fact? The most likely candidate is the non-burning of the forest in May. But is this an event? If you were sitting in the forest in May would you at some point say, "Look – the non-burning of the forest is occurring?" There are good arguments for saying that it is an event, and good arguments for saying that it isn't. But what's important here is that once this proposed counterexample is expressed in terms of event causation, the causal claims that it relies on lose some of their intuitive appeal: claiming that the April rain caused the non-burning of the forest in May sounds less appealing than claiming that the April caused there to be an unburnt forest in June. And intuitive appeal is all that the example has to rely on.

Second, the case of Dog-Bite. What does the dog-bite cause? "Me to press the button with my left forefinger" is what Lewis says. But again that is a fact. What event should we replace it with? The most likely is the one described by "the pressing of the button

[4] Lewis 1999, p. 31.

with my left forefinger". Now although it seems intuitively clear that the dog-bite caused me to press the button with my left forefinger, it is far from clear that the dog-bite caused the pressing of the button with my left forefinger. Although these may appear to make the same claim they are intuitively different, and they do not command assent and dissent equally. I, for one, am happy to agree with the first, but not with the second. (I think, furthermore, that I can explain my uneasiness: if the dog-bite had not occurred then my pressing the button with my left forefinger would not have been cleanly excised from the ensuing course of events – it would have been replaced by a similar event in which I press the button with my right forefinger. More of this later.) Again, the important point here is that when this example is expressed in terms of event causation, the causal claims involved lose some of their intuitive appeal.

Finally, the case of Shock C. This seems to be more resistant to the present line of defence. When given in terms of event causation the proposed counterexample goes something like this: A's moving his switch Right caused B's moving his switch Right, and B's moving his switch right caused C's shock, but A's moving his switch Right did not cause C's shock. The claim is that these are all intuitively correct and so transitivity fails. If there is a weakness here it is in the second claim: that B's moving his switch Right caused C's shock. I, for one, do not agree. I'm inclined to say that *the turning on of the switch* caused C's shock and that B's moving his switch Right merely enabled this causation to take place. But it can be argued that thinking of B's move as a mere enabler in this way will lead me into contradiction in other cases. Also, others apparently find it intuitive to say that B's move was a cause of C's shock. I think that since the example is meant to be nothing more than an appeal to intuition, and since it appeals to intuitions that some of us do not have, it cannot reasonably be seen as decisive against L2.

Whether or not Lewis's own defence is adequate, and whether or not these considerations are helpful to his case, his own position is that causation is invariably transitive and that L2 should not be rejected on the grounds that it isn't.

In the discussion above I touched on the issue of what it takes for an event to *not* occur: does it have to be neatly excised from the ensuing course of events, or can it be replaced by a similar event instead? Uncertainty over this issue raises problems for L2 in the following two cases:

> The Greeting. John greets Fred. Because he is tense, John says hello loudly. If he had not been tense, he would still have said hello, but softly. Fred jumps, and then returns John's greeting. If John had said hello softly, Fred would not have jumped, but he still would have returned John's greeting.
>
> The Vigorous Neuron. Neurons C_1 and C_2 fire, stimulating neuron B to fire vigorously, in turn stimulating neuron E to fire. If C_1 or C_2 had fired alone, then B would have fired feebly but would still have stimulated E to fire in the same way. If neither C_1 nor C_2 had fired, then neither B nor E would have fired.

Lewis grants that if L2 is to adequately deal with the first case, he must allow a *profligate* theory of events. Fred's return greeting needs to have a cause - Lewis does not want to say that it, or any event, is uncaused. If L2 is right, then it cannot be John's

saying hello loudly because there is no counterfactual dependence of the one on the other: if John hadn't said hello loudly, Lewis argues, then he might have said hello softly, in which case Fred would still have returned John's greeting. It must be, he says, the weaker event described by "John's saying hello" that causes Fred's reply, because only between these two events do we have the necessary counterfactual dependence: if John hadn't said hello then Fred would not have replied. We also need John's tension to have an effect – to say that it doesn't is to say that it makes no difference to the ensuing course of events, which is clearly wrong. If L2 is right, Lewis continues, then for this job John's saying hello will not do: if John had not been tense then he would still have said hello. Only John's saying hello loudly will do: if John had not been tense then we would not have said hello loudly. Thus, he concludes, in order for L2 to generate the appropriate causal statements we need to say that *both* events occurred – John's saying hello *and* John's saying hello loudly. Lewis accepts this profligacy, but it would certainly be better for L2 if he didn't have to.

His analysis of the second case is inconsistent with this. He thinks that L2 deems the firing of C_1 (and similarly the firing of C_2) as a cause of the firing of E, because there is a chain of counterfactual dependence from the first to the second: if the firing of C_1 had not occurred then the vigorous firing of B would not have occurred, and if the vigorous firing of B had not occurred then the firing of E would not have occurred. He says:

> My solution depends on assuming that if the intermediate event – the vigorous firing of B – had not occurred, then B would not have fired at all. It isn't that the vigorous firing would have been replaced by a feeble firing, differing only just enough not to be numerically the same.[5]

This is clearly inconsistent with his analysis of the first case. There he says that if John had not said hello loudly then he might have said hello softly. Here he says that if B had not fired vigorously then it would not have fired at all. Lewis could restore consistency in one of two ways. First, he could stick to what he says in the first case, and admit in the second that if the vigorous firing of B had not occurred then the feeble firing of B might still have occurred, and hence the firing of E might still have occurred as well. That would exclude B from any chain of counterfactual dependency from C_1 to E and from C_2 to E, so that L2 would deem neither a cause of E. Apart from leading to these arguably counterintuitive verdicts, this approach leaves us with the problem of profligacy in the first case. Second, he could stick to what he says in the second case, and say in the first that if John's saying hello loudly had not occurred then he wouldn't have said hello at all. Then John's saying hello loudly would do as both the effect of John's tension and the cause of Fred's return greeting, and so he could do away with the profligacy of events. It would also maintain the more intuitive result in the second that both C_1 and C_2 are causes of the firing of E.

Despite the obvious benefits of saying the second, McDermott has argued[6] that the right thing to say is the first, and that because L2 requires profligacy in The Greeting and counterintuitive verdicts in The Vigorous Neuron it must be inadequate. He claims that this is forced upon Lewis by the common-sense truth of these two counterfactuals:

(1) If John hadn't said hello loudly, he still might have said hello softly

[5] Lewis 1986, pp. 210-11.
[6] McDermott 1995a.

and

(2) If B hadn't fired vigorously, it still might have fired feebly.

I agree with McDermott that these are common-sensibly true. But I don't agree that they force Lewis into taking the position that McDermott claims. The reason is that these two statements express a counterfactual relation between *facts*, and L2 only makes a claim about counterfactual relations between *events*. For McDermott's argument to work, he needs to claim that *these* two counterfactuals are common-sensibly true:

(1') If John's saying hello loudly had not occurred, then John's saying hello softly might have occurred

and

(2') If B's vigorous firing had not occurred, then B's feeble firing might have occurred.

I maintain that they are not. I maintain that (1) and (1') have different truth conditions because their antecedents have different truth conditions, and similarly for (2) and (2'). When we entertain, in (1), the possibility that John hadn't said hello loudly, we seem to allow that John might have said hello in a different way instead (e.g. softly). The form of the words that we use suggests that the issue is not whether or not John says hello, but the way in which he says it. Not so for (1'). When we entertain the possibility that John's saying hello loudly did not occur, we seem to think of the event being completely and neatly excised from history, so that his saying hello softly is excluded as well. It is not the way in which he says hello that is at issue in this case, but whether or not he says hello at all.

McDermott claims that Lewis is forced in restoring consistency between his two analyses by adopting the first of the two strategies that we considered, and that since this leads to profligacy in the case of The Greeting and to counterintuitive verdicts in the case of The Vigorous Neuron L2 must be inadequate. Against this, I claim that Lewis is not so forced, and that by adopting the second strategy instead he can avoid these two problems and maintain that L2 is adequate.

In the case of The vigorous neuron, it seems intuitive to think that each of C_1 and C_2 is a cause of the firing if E. Although whether both C_1 and C_2 fired or only one of them has no effect on the firing of E, it does have an effect on the firing of B – B fires vigorously in the former case but only feebly in the latter. Such an event is called a *Bunzl* event. It is the occurrence of this Bunzl event that allows L2 (when applied in the way that I have argued it should be) to give the intuitively correct verdicts about C_1 and C_2: If the firing of C_1 had not occurred then the Bunzl event would not have occurred and if the Bunzl event had not occurred then the firing of E would not have occurred. So there is a chain of causal dependence from C_1 to E and C_1 is a cause of E (and similarly for C_2).

Actually, I need to be careful here. I claimed that the counterfactual, "If the firing of C_1 had not occurred then the Bunzl event would not have occurred", is true. I have argued that it is a feature of the counterfactual construction that the antecedent is true if and only if the firing of C_1 is completely and neatly excised from history. If I make the same claim about the consequent - that it is true if and only if the Bunzl event is completely and neatly excised from history - then the counterfactual comes out *false*. Why? Because if the firing of C_1 had not occurred then the Bunzl event would *not* have been

completely and neatly excised from history - it would have been replaced by a feeble firing instead. So if I want the counterfactual to come out true, then I have to say that the counterfactual construction is such that only the event in the antecedent needs to be completely and cleanly excised from history, not the event in the consequent (although it may be). That's what I will say. (Reluctantly, though. I don't like the asymmetry, and I have a feeling that things will work better in the long run if we avoid it. But here I am just trying to help Lewis with a particular problem.)

But what about the following case, in which there is no Bunzl event?

> Three Neurons. Neurons C_1 and C_2 fire, directly stimulating neuron E to fire. If either C_1 or C_2 had fired alone, E would still have fired in just the way it actually did fire. E would not have fired if neither C_1 nor C_2 had fired.

Since the firing of E is not counterfactually dependent on the firing of C_1 and there is no intermediate event to give a chain of counterfactual dependence (unlike in the case of The Vigorous Neuron), L2 deems that the firing of C_1 is not a cause of the firing of E. Similarly, it deems that the firing of C_2 is not a cause either. Two apparent problems for the results of L2 here: First, they make it sound as though E was uncaused - if neither C_1 nor C_2 was a cause then what was? Second, they seem counterintuitive – isn't it more natural to think of *both* the firing of C_1 and the firing of C_2 as a cause of the firing of E, rather than *neither*?

In response to the first, Lewis says that L2 does *not* deem the firing of E to be uncaused, because it deems the larger event consisting of the *mereological sum* (not the disjunction) of the firing of C_1 and the firing of C_2 to be a cause. He argues that if that event had not occurred (if it were completely absent, not replaced by a similar event), then the firing of E would not have occurred.

McDermott argues[7] that Lewis gets into trouble with this response. To get the result that the larger event caused the firing of E, Lewis has to appeal, he says, to the truth of this counterfactual:

> (3) If the larger event had not occurred, then the firing of E would not have occurred.

That is, McDermott claims, he has to appeal to the truth of this counterfactual:

> (4) If C_1 and C_2 had not both fired, then neither would have fired.

McDermott then says that not only is (4) contrary to common sense, it also implies, according to L2, that the firing of C_1 caused the firing of C_2: If C_1 had not fired then C_1 and C_2 would not both have fired, and hence, by (4), C_2 would not have fired (this line of reasoning assumes counterfactual transitivity, which does not invariably hold. But it can be checked that in this case it does).

Anyone who understands my response to McDermott above should be able to anticipate my response here. I agree with what he says about (4), but it is a counterfactual that

[7] Op. Cit.

relates *facts*, and so is *not* the counterfactual to which Lewis must appeal to get (3). The counterfactual that Lewis needs is this:

(4') If the firing of C_1 and C_2 had not occurred, then the firing of C_1 would not have occurred and the firing of C_2 would not have occurred.

I claim that this has different truth conditions to (4), for the same reason that (1) has different truth conditions to (1') and that (2) has different truth conditions to (2'). When we entertain the possibility of the firing of C_1 and C_2 not occurring, we imagine that the whole event – the firing of both of them – is completely and cleanly removed from history, in which case it is *true* that the firing of C_1 would not have occurred and the firing of C_2 would not have occurred. Moreover, it does not follow from (4') that if the firing of C_1 had not occurred then the firing of C_2 would not have occurred – although it *is* true that if the firing of C_1 had not occurred then C_1 and C_2 would not have both fired, it is *not* true that if the firing of C_1 had not occurred then the firing of C_1 and C_2 would not have occurred, because the firing of C_2 might have occurred, in which case the firing of C_1 and C_2 is not completely and cleanly erased from history. I think, then, that McDermott is wrong and that Lewis is right to think that L2 deems the larger event a cause of the firing of E.

Now to the second problem: that L2 seems to go against intuition when it declares that neither C_1 nor C_2 (individually) is a cause of E. Lewis's response is that this is no problem, because in cases like this we have no clear intuitions. But even if that is so, it still counts as a mark against L2 that it be decisive in a case in which intuition is not. Lewis would be better placed if he could argue that it *is* intuitively clear that the firing of C_1 is not a cause of the firing of E (and similarly for C_2). Or, if he can't do that, he would be better placed of he could argue that it is *wrong* to think that the firing of C_1 (or C_2) is a cause of the firing of E (i.e. that it is a misapplication of our concept of causation). I think that this second approach has some promise. Suppose we claim that C_1 (and hence, by symmetry, also C_2) is a cause of the firing of E. Suppose we remove C_1 from the situation so that we have only C_2 and E. C_2 fires and stimulates E to fire. Clearly we would want to say that the firing of C_2 is a cause of the firing of E. Now let's put C_1 back in to restore the original situation: C_1 and C_2 both fire and E fires, *exactly* as it did before. Can we *really* think that the firing of C_1 is now *also* a cause of the firing of E? Remember: C_2 still fires, the firing of C_2 is still a cause (we are granting) of the firing of E, and the firing of E occurs in exactly the same way as before? Can we *really* think that the firing of C_1 is a cause of the firing of E, even though it makes no difference to the firing of E? The answer might be: we were happy to do so in the case of The vigorous neuron, so why not here? Well, in that case we had a Bunzl event (the firing of B). Even though the firing of E occurs in exactly the same way whether or not C_1 fires, the firing of C_1 at least makes *some* difference – not to the firing of E, but to the firing of B. In the present case, what difference does C_1 make? Whether or not C_1 fires, *everything else* in the situation occurs in exactly the same way. Can we really think that an event is a cause even though it makes no difference?

There is probably no easy answer to this question. But I think it shows that we should be wary of accepting Three neurons as a genuine test case for L2 (or for any of Lewis's analyses). Anyhow, Lewis is not troubled by it, at least not as much as he is by this one:

> The Rock. Billy and Suzy throw rocks at a bottle. Suzy's rock arrives first. The bottle shatters. If Suzy's rock hadn't shattered the bottle then Billy's rock would have done so moments later.

Why is this a problem for L2? Intuitively, we are inclined to say that Suzy's throw was a cause of the bottle's shattering. So if L2 is right then we should also be inclined to say that there is a chain of counterfactual dependence from Suzy's throw to the bottle's shattering. Is there a one-step chain? (That is, is the bottle's shattering counterfactually dependent on Suzy's throw?) It seems not – if Suzy's throw had not occurred then the bottle's shattering would still have occurred (at the hand of Billy's rock instead). Is there a two-step chain? That would require there to be an event D (distinct from Suzy's throw and the bottle's shattering) such that if Suzy's throw had not occurred then D would not have occurred, and if D had not occurred then the bottle's shattering would not have occurred.

Finding such an event was easy in the case of The Push (e.g. Jones's collision with the truck) and The Bell (e.g. the ringing of bell A). Indeed, L2 was tailor made for cases like those. But here there seems to be a problem. The difficulty is not in finding an event to satisfy the first condition: the impact of Suzy's rock on the bottle will do – if Suzy's throw had not occurred then the impact would certainly not have occurred. The difficulty is in finding an event to satisfy the second as well. The impact of Suzy's rock on the bottle will not do – if it had not occurred then the bottle's shattering would still have occurred (from the impact of Billy's rock instead) (remember: no backtracking). What we need is an event between Suzy's throw and the bottle's shattering, distinct from both, that occurs after the back-up process (the one starting with Billy's throw) is cut short. But there is no such event, because this backup process is cut short only by the shattering of the bottle itself. So there is no 2-step chain of causal dependence from Suzy's throw to the bottle's shattering, and, by similar reasoning, no n-step chain for any n>2. According to L2, and contrary to intuition, Suzy's throw is *not* a cause of the bottle's shattering.

Lewis calls cases like The Rock cases of *late preemption*, in contrast with cases like The Push and The Bell which he calls cases of *early preemption* (because their backup processes are cut short early enough to provide the intermediate events that L2 needs). It seems that the Rock (and every other case of late preemption) is a counterexample to L2. Lewis accepts it as such and modifies his analysis accordingly. But before we look at that, there is one line of defence that we should consider.

It could be argued that there is indeed a one-step chain of causal dependence from Suzy's throw to the bottle's shattering, that is, that the latter is directly counterfactually dependent on the former. If Suzy's throw had not occurred, the argument runs, then the rock's shattering (the one that actually occurred) would *not* have occurred, because the shattering at the hand of Billy's rock would have been a *different* shattering, a different event – it would, for instance, have happened slightly later. This is a similar objection to the one we considered in the cases of The Push and The Rock. Lewis's response is that to argue in this way is to impose unrealistic fragility conditions on the shattering of the rock. He says:

> 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 cancelling it because I wanted *that very event* to take place.[8]

Because, as Lewis points out, there are some occasions in which we think that the same event can take place at quite different times and in quite different ways, anyone who wants to appeal to the fragility of the bottle's shattering owes us an account of *why* this particular event is fragile. The natural (and possibly only) answer is that the other shattering would have happened for a different reason, would have had a different cause, would have been the result of a different causal path. But he cannot rely of causal notions in this way without giving up the possibility of L2 providing a *reductive* analysis of causation.

To summarise, we have looked at three ways of arguing that L2 is inadequate. The first was to note that it implies the invariable transitivity of causation and that Forest Fire, Dog-Bite and Shock C seem to be cases in which this transitivity fails. We saw that Lewis rejects this as a problem for L2, because he thinks that in each case transitivity *does* succeed. We also saw a way of arguing that once these cases are presented in the language of event causation, as they ought to be, then they lose some of their intuitive appeal. The second was to argue that L2 fails in cases like The Greeting, The Vigorous Neuron, and Three Neurons, because the non-occurrence of some events can happen in more then one way. We saw that Lewis rejects this as a problem for L2 as well, but not without some difficulty. We offered an alternative reason to think that these cases are successfully handled by L2. The third was to argue that L2 fails in cases of late preemption like The Rock, because the backup process is cut short so late that there is no chain of counterfactual dependence from what we want to say is the cause to the effect. Lewis accepts this as a genuine problem for L2 and offers in its place the analysis that we will consider next.

5. The Third Analysis: Quasi-dependence

We have seen that cases like The Push and The Bell suggest that causation is transitive in a way that counterfactual dependence is not, and that to allow for this Lewis generalised his analysis of causation from being direct counterfactual dependence to being chains of counterfactual dependence instead. But we have also seen that the new analysis seems to fail in cases like The Rock, where we have causation but no such chain. Lewis's next analysis is based on the idea that causation is *intrinsic* in a way that counterfactual dependence is not, and that allowance needs to be made for this fact as well.

Why might Lewis think that causation is somehow intrinsic? In the case of The Rock, we are inclined to think that Suzy's throw is a cause of the bottle's shattering, whether or not Billy happens to throw his rock as well. Our idea seems to be this. Say that a *process* is any course of events. If only Suzy were to throw her rock then there would be a process, P, starting with Suzy's throw and ending with the shattering of the bottle, in virtue of which we would say that Suzy's throw is a cause of the shattering. Say that in this case P is a *causal process* from Suzy's throw to the shattering. When Suzy and Billy both throw P still takes place – that is, all of the events in P still occur. Moreover, we still think that P is a causal process, despite the occurrence of new events and new

[8] Lewis 1999, p.15.

processes. We seem to think that whether or not P is a causal process does not depend upon the occurrence (or non-occurrence) of events not included in P. That's what Lewis means by causation being intrinsic. The problem for L2 is that chains of causal dependence do not behave like this. If only Suzy were to throw, then there is a chain of causal dependence from Suzy's throw to the bottle's shattering (in fact, a one-step chain). But if Billy throws as well then there is no such chain, despite the fact that the process of events from Suzy's throw to the bottle's shattering remains unaffected. This is the motivation behind his next analysis – the need to make chains of causal dependence intrinsic in the way that causation seems to be.

To this end, he says that event E *quasi-depends* on event C if and only if some process that has C as its first member and E as its last member has an intrinsic duplicate in the same world, or in some other world with the same laws, such that its last member counterfactually depends on its first.[9] Then he says that event E *causally depends* on event C if and only if it either counterfactually depends on C or quasi-depends on C. Then:

> L3: C is a cause of E if and only if there is a chain of causal dependence from C to E.

How is this meant to work in the case of The Rock? Consider the process that starts with Suzy's throw, includes the events that constitute the flight of her rock, and ends with the shattering of the bottle. There is a possible world which has an intrinsic duplicate of this process but in which Billy and his rock are entirely absent. In this comparison case the shattering of the bottle is counterfactually dependent on Suzy's throw, so in the actual case the shattering of the bottle is quasi-dependent on Suzy's throw, so there is a one-step chain of causal dependence from the former to the latter, so the former counts as a cause of the latter. What about Billy's throw? L3 *does* correctly declare that it is *not* a cause of the shattering, but it takes a bit more work to see why. Firstly, the shattering does not counterfactually depend on Billy's throw – if Billy's throw had not occurred then the shattering would still have occurred. Secondly, it does not quasi-depend on Billy's throw either. For consider any process that starts with Billy's throw and ends with the shattering of the bottle, and consider any intrinsic duplicate of this process in the actual world or in some possible world with the same laws. In the actual process there is no event of Billy's rock hitting the unshattered bottle (because Suzy's rock shattered the bottle before Billy's rock arrived). But it follows from the laws of the actual world that bottle's do not shatter unless something hits them. So in keeping with these laws, in the comparison case there must be an event of *something* hitting the unshattered bottle. It can't be Billy's rock because that would take away its intrinsic similarity with the actual process. It must be something else. But then the shattering does not counterfactually depend on Billy's throw – if Billy's throw had not occurred then the bottle's shattering would still have occurred, because this other thing would still have hit the bottle. This is true of *all* intrinsic duplicates of *all*

[9] At least, this is how he defines it in Lewis 1999, p. 11. Earlier (Lewis 1986, p. 206) he said that for E to quasi-depend on C *the great majority* of comparison processes, as measured by variety of the surroundings, must exhibit the proper pattern of dependence, not just one. I think that this is unnecessarily complicated and problematic, so I have gone with his more recent definition.

processes starting with Billy's throw and ending with the shattering of the bottle. So the latter does not quasi-depend on the former. Thirdly, since the shattering is neither counterfactually dependent nor quasi-dependent on Billy's throw it is not causally dependent on it either. So there is no one-step chain of causal dependence. Fourthly, there is no two-step chain of causal dependence. To get such a chain we need some event D, distinct from Billy's throw and the bottle's shattering, such that D counterfactually depends or quasi-depends on Billy's throw and the bottle's shattering counterfactually depends or quasi-depends on D. We know from our failed application of L2 to this case that there is no such event on which the bottle's shattering *counterfactually* depends. Is there one on which it *quasi*-depends? The answer is no, for the same reason that the bottle's shattering does not quasi-depend on Billy's throw: in any comparison case there must be something other than Billy's rock that shatters the bottle, and its presence breaks the counterfactual dependence that we need. Finally, by similar reasoning we can see that there is no n-step chain of causal dependence, for any $n > 2$.

Lewis's definition of quasi-dependence relies on the poorly defined notion of being an "intrinsic duplicate", and so this application of L3 is certainly open to objection. It is not clear, for example, that because any process from Billy's throw to the bottle's shattering does not include the impact of Billy's rock on the unshattered bottle that any world that contains an intrinsic duplicate of this process cannot also contain such an event (a claim that was appealed to above). But we will not concern ourselves with these problems for L3. It and the notion of quasi-dependence face deeper problems, ones that Lewis accepts as fatal and lead him to reject the analysis. In particular, Lewis accepts the three cases to follow as genuine counterexamples to L3.

First, a modification of The Rock:

> Jumping Rocks. Billy and Suzy throw identical rocks at a bottle, in a world whose laws allow flying rocks to make tiny (deterministic) jumps. Suzy's rock arrives first and shatters the bottle. No jumps occur.[10]

Why does Lewis accept this as a counterexample? Even though the laws allow them no jumps actually occur, so we are still inclined to say that Suzy's throw and not Billy's was a cause of the bottle's shattering. But according to L3 we should be inclined to say that Billy's is a cause as well. For consider the process starting with Billy's throw, including the flight of Billy's rock up to some time a little before its arrival at the shattered bottle, and finishing with the impact of Suzy's rock on the bottle and the bottle's shattering. There is a world with the same laws in which Suzy and her throw are entirely absent, and in which we have a process starting with Billy's throw, including the flight of Billy's rock up to the same time and then a tiny jump of Billy's rock, and finishing with the impact of Billy's rock on the bottle and the bottle's shattering. Because Suzy's and Billy's rocks are identical, the original and comparison processes are intrinsic duplicates, and in the comparison case the shattering is counterfactually dependent on Billy's throw. So in the original case it is quasi-dependent on Billy's throw and the latter is, counterintuitively, a cause of the former.

[10] In the case that Lewis considers, the laws of nature allow that rocks make tiny *random* jumps. I have said that the laws allow tiny *deterministic* jumps, in keeping with my assumption of determinism. I think that everything goes through just the same.

Second:

> The Sergeant and Major. The soldiers know that they must obey the order of the most senior officer. The Sergeant and Major simultaneously shout 'Advance!'. The soldiers advance.

Lewis thinks that this is, intuitively, a case of preemption: the Major is a cause of the soldiers's advance; the Sergeant is not, but would have been had the major said nothing. Moreover, it is a special type of preemption, different to any case of preemption we have seen so far. In each of those cases, the preemption is brought about by *cutting*: the backup process does not run to completion, but is cut short by the causal process, has some of its events prevented by the causal process. In The Push, the causal process cuts the backup process by preventing the collision of Jones with the bus; in The Bell, the ringing of bell B is prevented; in The Rock and Jumping Rocks, the impact of Billy's rock with the unshattered bottle is prevented. But in the present case, it is claimed, there is no cutting – the backup process runs to completion: the Sergeant shouts 'Advance!', the soldiers hear him, decide to advance, and advance. This kind of case is called *trumping preemption*.

Trumping preemption is a problem for L3. Although it correctly deems that Suzy's throw is and Billy's throw is not a cause of the bottle's shattering, its success relies on the backup process being cut short: it is because the impact of Billy's rock on the unshattered bottle is missing in the actual case that no process from Billy's throwing to the bottle's shattering has an intrinsic duplicate that is a genuine causal process. But in cases of trumping preemption the backup process is not cut short, so L3 incorrectly deems it to be a genuine causal process. Let's consider how this happens in the case of The Sergeant and Major. Take the actual process (Sergeant shouts 'Advance!', soldiers hear sergeant, soldiers decide to advance, soldiers advance). This process has an intrinsic duplicate in a world in which the Major is absent, and the last event in this duplicate is causally dependent on the first. So in the actual world the soldiers's advance is quasi-dependent on the Sergeant's order, and the latter is deemed to be a cause of the former.

The defender of L3 can, of course, object to this. At least these two things can be said: (i) intuitively the Sergeant's order *is* a cause of the soldiers's advance, so the result obtained by this application of L3 is the right one, or (ii) this is just another case of cutting, and if L3 is correctly applied then it will *not* deem the Sergeant's order to be a cause.

It might seem that the first objection is easy to dismiss. Think about *why* a particular soldier advances: he hears the Sergeant and Major both shout, he knows that he must obey the Major, so he ignores the Sergeant and does what the Major says – he advances because the Major shouted 'Advance!', regardless of what the Sergeant happened to shout. When described in this way, it seems intuitive to think that the Major's order was a cause of the soldier's advance and that the Sergeant's was not. But, the objection counters, we could also think about it this way: the soldier hears the Sergeant and Major shout the same thing, he notes that the Major has not overruled the Sergeant so that he can ignore the Major's order and do what the Sergeant says – he advances because the Sergeant shouted 'Advance!'. When described in this way, the objection suggests, it seems intuitive to think that the Sergeant's order was a cause of his advance and not the

Major's. Thus, the objection concludes, it is wrong to claim that the Sergeant's order is not a cause of the advance.

This is a bad line of defence, for three reasons. First, it loses sight of the fact that all we need for this to be a counterexample is for it to be *possible* to think, consistently with the described course of events, that the Major's order is a cause of the advance and that the Sergeant's is not, which it certainly seems to be. It *may* also be possible to think, in the way that I described, that this is the wrong way around – that the Sergeant's order is a cause and the Major's is not. But that is not to the point. Second, even this alternative way of thinking about the situation is a problem for L3 because it suggests, contrary to L3, that the Major's order is not a cause of the soldier's advance. What the defender would need to run this line of argument is a way of thinking about the situation that makes it intuitive to think that *both* the Sergeant's and Major's orders are causes, and that seems to be hard to find – it is difficult to think of the soldier following both orders, rather than following one instead of the other. Third, there seems to be an asymmetry between the Sergeant's order and the Major's order that shows up in these two ways of thinking. In the first, we can say that the soldier advances because the Major shouted 'Advance!'*, regardless of what the Sergeant happened to shout*. In the second, we are reluctant to make the corresponding claim - that the soldier advances because the Sergeant shouted 'Advance!', *regardless of what the Major happened to shout* - because if the Major had shouted something different then the soldier would not have advanced. Against this, it could be said that even though the soldier's decision to do what the Sergeant shouted relied on the fact that the Major shouted the same thing, once he had made that decision he did do what the Sergeant said regardless of what it was that the Major said. Even if this response is adequate, however, it does not take away the feeling that there is an asymmetry between the two orders, an asymmetry that is not captured by L3.

What about the second objection, that this is just another case of cutting so that if we were to apply L3 correctly then the Sergeant's order would not come out as a cause? To be an adequate line of defence, the claim needs to be not that we *do* think that the Sergeant's order is pre-empted by cutting, but that we *must* think it is pre-empted by cutting. Why? Because all we need for the case to be a counterexample to L3 is for it to be possible to think, consistently with the described course of events, that the Major's order is a cause of the advance and the Sergeant's is not, while thinking that the Sergeant's backup process is not cut. It might seem wrong to claim that we *must* think the Sergeant's process is cut – if we think about the situation in the second way described above it seems natural to think that it's not: the soldiers ignore what the Major says and continue to obey the Sergeant's order. But if we *do* think about it that way, we also think that the Sergeant's order is a cause and not the Major's. Can we think that the Sergeant's process is not cut and at the same time think that the Major's order is a cause and that the Sergeant's is not? I think it's difficult to argue "yes", and that this is a good line of defence.

Note that both of these objections rely on there being extra events between the Sergeant's and Major's orders and the advance (specifically, mental events in the soldiers's heads). The next example is proposed as a way of blocking them at the outset, by stipulating that there are no such events:

> The Wizards. The laws of magic state that what will happen at midnight must match the first spell cast that day. As it happens, Merlin casts a 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. No other events occur.

It is claimed by those who offer this example that it presents a similar difficulty for L3 as does The Sergeant and Major: our intuitive judgment is that Merlin's spell is a cause of the prince-to-frog transformation and Morgana's is not, and yet according to L3 they both are (the application if L3 is very similar in each case). Furthermore, because there are no events other than the two spells and the transformation, there is no story about what else occurs to make it plausible that Morgana's spell was indeed a cause, nor can it be objected that this is just another case of cutting.

It is hard to deny that if we accept these intuitive verdicts then Merlin's spell pre-empts Morgana's without cutting: The back-up process (Morgana's spell, transformation) runs to completion whether or not Merlin casts his spell. So the defender of L3 must resist the intuitive verdicts. Should we accept them? Recall that this is what we are being asked to accept: Merlin casts a spell in the morning, Morgana casts a spell in the afternoon, a transformation takes place at midnight, no other events occur, Merlin's spell is a cause of the transformation, Morgana's spell is not. Taken individually, it seems reasonable to accept each of these claims, but I think we must be careful about accepting them *all together*. In particular, we must be careful about accepting that Merlin's spell is a cause of the transformation and Morgana's is not, even though no other events occur. To emphasise, *no other events occur* – the two wizards cast their spells, and later the transformation occurs. There is no transfer of energy from wizards to prince, there is no chief wizard keeping tabs on which spells are cast, there are no other events at all. Is this a situation in which our concept of causation *genuinely* applies? Can we *really* think that causation can take place this way? Sure, we may be happy to accept that there is causation here, albeit a kind of causation that we are not used to. But do we honestly accept that nothing else occurs, or do we believe that if we looked closely enough we would find some more events? I think that this is a difficult issue, and one that needs some careful thought before deciding whether or not The Wizards is a genuine problem for L3. But whether or not Lewis has given the matter such thought, he accepts that it is.

To summarise. We have seen that Lewis introduced the notion of quasi-dependence to allow for the fact that causation is an intrinsic relation. The new analysis seems to solve the problem faced by its predecessor in cases of late preemption like The Rock. But it is not so successful in cases like Jumping Rocks, where the laws of nature allow rocks to be not-so-well behaved, and in cases of trumping preemption like The Sergeant and Major and The Wizards, where the backup process is pre-empted without cutting. We have seen ways to argue that these last two do not pose a genuine problem, but Lewis accepts that they do and offers the analysis that we will look at next.

6. The Fourth Analysis: Influence

The motivating idea behind quasi-dependence was that causation is an intrinsic relation between events, except in so far as being subject to the laws of nature is an extrinsic matter. Lewis accepts that the case of The Wizards shows this to be "an over-hasty generalization" about the causation that happens in other worlds: whether or not

Morgana's spell is a cause of the transformation depends on more than just the laws and the intrinsic nature of the (Morgana's spell, transformation) process – it also depends on whether or not Merlin casts an earlier spell. Moreover, he accepts that the case of The Sergeant and Major may even show it to be an over-hasty generalisation about the causation that happens in our own world. He says, "It may be, for all we know, that our case of the soldiers obeying the Major is a trumping case that *actually* happens"[11] (my emphasis). He also rejects quasi-dependence independently of the occurrence of trumping preemption. He says:

> 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.[12]

What Lewis needs is an analysis which captures *in some other way* the intuitive asymmetry between Suzy's throw and Billy's throw, between the Major's order and the Sergeant's order, and between Merlin's spell and Morgana's spell. To this end, he turns to the idea of *influence*…

Let the *fragility* of an event be the extent to which it can occur in only one place, time and manner. Let an *alteration* of an actual event E be either a very fragile *version* of E (numerically the same but otherwise possibly different) or else a very fragile *alternative* event (numerically different, but otherwise possibly very similar). One alteration of E is the alteration that actually occurs – call this the *unaltered* alteration. The other alterations are all unactualised. Leave it unspecified what it takes for an alteration of E to be an alternative rather than a version – if we think that E is very fragile then we will think that all of its unactualised alterations are alternatives, numerically different from E itself. If we think that E is not at all fragile then we will think that all of its alterations are different versions of one and the same event. Or we might think that some are alternatives and some are versions, or we might have no opinion. Where C and E are distinct actual events, say that C *influences* E if and only if there is a substantial range C_1, C_2, … of different not-too-distant alterations of C (including the actual alteration of C) and there is a range E_1, E_2, … of alterations of E, at least some of which differ, such that if C_1 had occurred then E_1 would have occurred, and if C_2 had occurred than E_2 would have occurred, and so on. Then say:

> L4: C is a cause of E if and only if there is a chain of influence from C to E.

This is how Lewis describes his motivating idea:

> 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

[11] Lewis 1999, p. 12.
[12] Loc. Cit.

> 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.[13]

L4 is a significant change from the previous three analyses, so before we see how it deals with cases of late preemption we had better see how it deals with our earlier ones. The Word, Socrates Death, and The Thermometer are no problem because the requirements (i) that C and E be distinct, and (ii) that backtracking counterfactuals are not allowed, are both kept. The Push and The Bell are no problem: My pushing Jones has influence over his death because had I not pushed then his death at the hands of the bus would have been a different alteration of his death at the hands of the truck, so L4 says we should think that my pushing Jones is a cause of his death, as we do; Tom's firing of the bullet influences Harry's jumping with fright, because had Tom fired a little sooner or later Harry would have jumped a little sooner or later, so L4 says we should think that Tom's firing is a cause of Harry's jumping, as we do. Note that chains of dependence were added to the earlier analyses in order to correctly deal with these two cases - cases in which transitivity holds for causation but not for causal dependence. But L4 gives the intuitively correct verdicts using only direct influence (i.e. without resorting to chains of influence of more than one step). Lewis considers the possibility that influence is always transitive and that chains of influence can be done away with. That turns out, however, not to be true. See Lewis 1999, p. 25-28, for details. Finally, since L4 does allow chains of influence it, like L2 and L3, is committed to causation being invariably transitive, and so faces refutation at the hands of Forest Fire, Dog-Bite and Shock C to just the same extent.

Now, how does L4 deal with our three cases of late preemption? First, how does it deal with The Rock? Suzy's throw influences the bottle's shattering, because had she thrown a little sooner the bottle would have shattered a little earlier, and had she aimed at the neck instead of the side the bottle would have shattered in this way rather than that, and so on. So there is a one-step chain of influence from Suzy's throw to the shattering, and L4 declares the former to be a cause of the latter. But Billy's throw influences the bottle's shattering as well, because had he thrown a little sooner its gravitational effects on Suzy's rock would have been slightly different and so Suzy's rock would have followed a slightly different trajectory and the bottle's shattering would have been just a bit different, and similarly if he had aimed at the neck instead of the side, and so on. So it seems that L4 incorrectly declares that Billy's throw is also a cause. Lewis accepts that it *does* declare Billy's throw to be a cause, but does not accept that this is a mistake. He points out that it declares Suzy's throw to be *much more* of a cause: "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"[14], and he seems to think that this agrees with intuition, that it is intuitively correct to say "Suzy's throw is *much more of a cause* of the bottle's shattering than Billy's"[15] (my emphasis). It may seem that by altering Billy's throw so that his rock arrives first we

[13] Lewis 1999, p. 22.
[14] Lewis 1999, p. 24.
[15] Loc. Cit.

can make a lot of difference to the shattering of the rock – just as much difference, in fact, as we can by altering Suzy's throw. But L4 says that we only consider not-too-distant alterations of Billy's throw when thinking about what influence it has, and presumably we think that the alterations in which Billy's rock arrives first are too-distant. There is a vagueness built in to L4 here – it is left unclear which alterations of C count as "not-too-distant". Lewis hopes that this vagueness matches the vagueness of our concept of causation and hence need not be regretted. I wonder why, then, he does not also build in vagueness about which alterations of E are *distant-enough*? L4 says that when deciding if C influences E we only consider not-too-distant alterations of C, which seems to be right. But we also seem to ignore alterations of E that are not-distant-enough – we want to say, "of course alterations to Billy's throw make *some* difference to the shattering, but they are so small that they should be ignored". So why not include this in L4 as well? Why not stipulate that the alterations of E be distant-enough? Then Lewis can get the more intuitive result that Billy's throw is not a cause at all, instead of the less intuitive one that it is a cause, but much less of one.

How does L4 handle The Sergeant and Major? The Major's order has influence over the soldier's advance: if the Major had shouted 'Take cover!' the soldiers would have taken cover, if he had shouted 'Retreat!' they would have retreated, and so on. So L4 pronounces The Major's order to be a cause of the soldiers's retreat. (Presumably, we think that the alterations in which the Major shouts 'Take cover!' and 'Retreat!' are not-too-distant.) The Sergeant's order, on the other hand, has much less influence: no matter what the Sergeant had shouted, the soldiers would have advanced at almost exactly the same time and in almost exactly the same way (almost, because the difference in the sound waves produced by the Sergeant would no doubt have had *some* effect on the advance). (Presumably, any alteration in which the Sergeant's order comes early enough to be enacted before the Major has a chance to trump it is thought to be too-distant.) Again, we have the result that the Sergeant's order is a cause, but much less of one than the Major's order.

How does L4 handle The Wizards? Merlin's spell influences the transformation: had it been king-to-kangaroo rather than prince-to-frog then the transformation would have been correspondingly different. But Morgana's spell does not: had it been something other than prince-to-frog then the transformation at midnight would have been no different. (Presumably, again, any alteration of Morgana's spell in which it is the first one cast is thought to be too-distant.) Whereas the Sergeant's order in the case above has *some* influence, Morgana's spell seems to have *no* influence: the transformation would have been *exactly* the same regardless of what Morgana had said. So in this case, L4 is not forced into saying that Morgana's spell is a cause, albeit much less of one.

There are at least three problems for L4. The first is that it seems to vastly overgenerate. In the case of The Rock it declares that Billy's throw is a cause of the bottle's shattering (much less of a cause than Suzy's throw, but still a cause), and in the case of The Sergeant and Major it declares that the Sergeant's order is a cause of the soldiers's advance. But we are inclined to think that these events are not causes *at all*, although they would have been had the potential causal processes they began not been pre-empted. In fact, L4 declares many events to be a cause of some event, E, that we would not normally think of as causes – any event that makes some difference to the time, location or manner of E (no matter how small). But Lewis seems to think that L4 is right in this – that all of these events *are*, to varying degrees, causes of E. Of the

asymmetry between Suzy's and Billy's throw he says, "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."[16]

That's the first problem. Next, note that L4 doesn't seem to overgenerate in the case of The Wizards – Merlin's spell influences the transformation so is declared a cause, Morgana's spell does not so is declared not a cause. So it seems that L4 gets it right at least in cases like these. But anyone who accepts The Wizards as a genuine test case for Lewis's analyses ought to accept the following case as well, and this is the second problem for L4:

> One-Spell Wizard. The laws of magic state that what will happen at midnight must match the first spell cast that day, and that only a prince-to-frog utterance counts as a spell for Merlin. Merlin casts his spell in the morning, Morgana casts another prince-to-frog spell in the evening. At midnight the prince turns into a frog.

It is claimed by those who offer this as a counterexample that it is still intuitive think that Merlin's spell was a cause of the transformation and Morgana's was not. This is a problem for L4, because Merlin has *no influence*. Altering *when* Merlin spoke makes no difference to the transformation at midnight – he and Morgana both cast a prince-to-frog spell, so no matter who spoke first the first spell that day the transformation at midnight would have been prince-to-frog. Altering *what* Merlin said makes no difference either. Suppose Merlin had uttered something other than he did. If it would have still counted as a prince-to-frog utterance then it would have still counted as a spell, it would have still been the first spell that day, and the transformation at midnight would still have been prince-to-frog. If it wouldn't have counted as a prince-to-frog utterance then it wouldn't have counted as a spell, and so Morgana's spell would have been the first that day. But hers was a prince-to-frog spell anyway, so the transformation at midnight would still have been prince-to-frog.

Lewis could defend L4 by rejecting the intuitions of those who offer One-Spell Wizard as a counterexample, by claiming that it is intuitively just like the case of Three Neurons - neither Merlin's spell nor Morgana's spell is a cause of the transformation, but their mereological sum is. If it *is* intuitive to think this, then One-Spell Wizard is no problem for L4, because these are exactly the verdicts that it delivers. But even if we grant Lewis this defence, there is a third problem:

> Wizard in Waiting. Merlin the one-spell wizard casts his one and only prince-to-frog spell in the morning. Morgana casts no spell. At midnight the prince turns into a frog. Had Merlin not cast his spell, Morgana would have cast a prince-to-frog spell in the evening.

Even if Lewis can successfully deny that Merlin's spell is intuitively a cause of the transformation in the previous case, it would be very hard for him to deny it here. After all, only two events occur - Merlin's spell, and the transformation. If Merlin's spell wasn't a cause of the transformation, then what was? But L4 still delivers the verdict that Merlin's spell is not a cause because it still has no influence (I'll leave it to the reader to check). I know of no good defence of L4 against this problem.

[16] Op. Cit., p. 20.

To summarise. We have seen that Lewis rejects quasi-dependence as a way of solving the problems forced upon L2 by cases of late preemption, and tries to do so instead by analysing causation in terms of influence. This latest analysis, L4, seems to work for The Rock, The Sergeant and Major, and The Wizards. But we noted that it faces at least three problems. First, it makes causation a matter of degree, and judges any event that has even the tiniest influence over E to be a cause of E. This seems very counterintuitive. Lewis simply disagrees. Second, it seems to counterintuitively judge Merlin's spell as not a cause of the transformation in the case of One-Spell Wizard. It might be open to Lewis to argue that this is a case, like Three Neurons, in which intuition is too unclear to be a reliable guide. But even if that is right, it seems intuitively clear in the case of Wizard in Waiting that Merlin's spell is a cause, and yet L4 says that it's not. I don't know what Lewis has to say about this third problem, and I have no help to offer.

7. Some Final Remarks

We have now seen Lewis's four most significant analyses, some of the problems that they face, and some of the issues involved in judging the extent to which they succeed. Intuitively, there does seem to be a close connection between causation and counterfactual dependence, at least part of which seems to be captured by each analysis. But, as I hope this essay shows, it is frustratingly difficult to work out exactly what that connection is. Lewis believes that counterfactual dependence, in some form or another, is constitutive of causation - that if we have full knowledge of the facts and counterfacts about a particular course of events then we have full knowledge of the causal facts as well. The challenge is to specify how the causal facts are thereby determined. But smart people keep dreaming up smart counterexamples - courses of events for which Lewis's analyses generate the *wrong* causal facts from the facts and counterfacts. I, for one, can't help but feel that no matter what analysis Lewis gives such counterexamples will always be available. The connection between causation and counterfactual dependence is close, but it is not, I suggest, close enough to guarantee that for *no* course of events do they come apart.

References

Lewis, D.

[1973] *Counterfactuals*, Oxford, Blackwell.

[1979] "Counterfactual Dependence and Time's Arrow", *Nous*, **13**, pp. 455-76.

[1986] *Philosophical Papers Vol. II*, Oxford, Oxford University Press.

[1999] "Causation as Influence", To appear.

McDermott, M.

[1995a] "Lewis on Causal Dependence", *Australasian Journal of Philosophy*, **73**, pp. 129-39.

[1995b] "Redundant Causation", *Brit. J. Phil. Sci.*, **46**, pp. 523-44.

Schaffer, J.

[1999] "Trumping Preemption", To appear.