# Transitivity and the Ontology of Causation

NICHOLAS UNWIN

**Work in progress seminar: 17th February, 2010**

ABSTRACT
It is argued that it is very hard to analyse causation in such a way that prevents everything from causing everything else. This is particularly true if we assume that the causal relation is transitive (i.e., if A causes B and B causes C, then A automatically causes C), for it all too often happens that causal chains that we wish to keep separate pass through common intermediate events. It is also argued that treating causes as aspects of events, rather than the events themselves, will not resolve this problem.

## 1 INTRODUCTION

It has seemed obvious to most philosophers that causation is a transitive relation. For example, if your existence is caused by that of your parents, and if their existence is caused by that of your grandparents, how could your grandparents' existence fail to cause yours? It is not a proximate cause, to be sure, but it is obviously a cause nonetheless. To doubt this seems perverse, and the point can surely be generalized.

Yet there are a number of well known apparent counterexamples.

**Case 1**
Jones plans to push a button to detonate the bomb. A dog then bites off his right forefinger, thus causing him to use his left finger instead (he is normally right-handed). The bomb duly goes off. This was caused by his pushing the button with his left finger, which was caused by the dog's biting off his right finger. But the latter surely did not cause the bomb to go off. The dog had nothing to do with it. (From McDermott 1995, cited by Hall 2004)

**Case 2**
A small fire (A) breaks out in a downstairs room of a large house. This immediately triggers an automatic sprinkler system (B), thanks to which the rest of the house is unscathed (C). Here, A clearly causes B, which in turn causes C. But, intuitively, A does not cause C. (My example)

Counterfactual or dependence theories of causation are immediately vulnerable to this problem, if only because the counterfactual conditional is notoriously non-transitive (though there are other problems as well). For example,

> If the automatic sprinkler had not operated, the whole house would have burnt down

and

> If there had not been a small fire in the downstairs room, then the automatic sprinkler would not have operated

do not jointly imply

> If there had not been a small fire in the downstairs room, the whole house would have burnt down

In general, the inference

$$P \mathrel{\square\!\!\rightarrow} Q, Q \mathrel{\square\!\!\rightarrow} R \vdash P \mathrel{\square\!\!\rightarrow} R$$

is invalid. David Lewis's analysis of counterfactuals (in terms of what goes on in the possible worlds most similar to ours) explains this, and related oddities, quite neatly.

How do we get past this? Some philosophers (e.g., Ned Hall) deny that counterfactual dependence is required for causation in this straightforward sort of way:

> ... the price of transitivity—a price well worth paying—is to give up on the claim that there is any *deep* connection between counterfactual dependence and (the central kind of) causation. (Hall 2004: 182, original emphasis)

Lewis, himself one of the best known defenders of counterfactual analyses of causation, originally defined causal dependence as the ancestral of the appropriate sort of counterfactual relation. Obviously, this automatically guarantees transitivity, and he never adjusted this aspect of this theory. In both cases, it will be agreed (I suppose) that the small fire actually did cause the whole house *not* to be burnt down. It is just that we lack direct, unmediated counterfactual dependence. This may sound odd, but it could be pointed out that a complete causal history of a given event is bound to contain elements whose connection is seriously indirect. We should therefore not have to worry too much about the unintuitive cases.

Others, such as L.A. Paul (2004), focus more on examples such as Case 1. Here, what goes wrong is that the dog's bite only caused one *aspect* of the bomber's pushing the button, namely that it was done with the left forefinger rather than the right; and *that* aspect was not causally relevant to the detonation. She concludes that causes and effects are not events as such, but rather aspects of events, which are instantiations of specific properties at various places and times. The trouble with events is that they are individuated too coarsely: exactly how coarsely they are individuated is, of course, controversial, but most would agree that Jones's pushing the button and his pushing it there and then with his left forefinger are not two different events. This problem ensures that, as a causal chain progresses through events, it gradually picks up more and more irrelevant aspects which in turn generate more and more irrelevant effects. Transitivity will thus fail unless we allow that virtually anything causes anything. However, if we focus on the aspects themselves instead of the 'fat' events that they are aspects of, then causal chains are far 'thinner'. They transmit only what is directly relevant, and our ordinary intuitions about dependence and transitivity can therefore (perhaps) be preserved.

It is often thought that the serious (apparent) counterexamples to causal transitivity all share a common feature, namely that they involve a certain kind of *double prevention*. Thus the small fire aims to prevent the survival of the house, but the automatic sprinkler provides a counter-prevention; likewise, the dog bite prevents Jones from pushing the button in the way he planned, but his switching to the left hand prevents that from preventing the ultimate effect. We thus end up with a kind of double-negative, so to speak, and the influence of the original cause is thus cancelled out (apparently). This paper will argue, however, that transitivity failures are far more widespread than might be thought, unless causes and effects are understood to be 'aspects' as understood in a very narrow sense indeed—far narrower than usually thought. Furthermore, it will be argued that, when our causal ontology is narrowed to this extent, other problems emerge which equally make it hard to stop anything from causing anything else.

## 2 'CAUSAL CROSSOVERS'

To see why we have a general problem, we need to consider a slightly different range of examples. To begin with, consider the following events:

*a* = my desire that the car not stall
*b* = my car's travelling at more than 5 mph
*c* = my car's not stalling

*d* = my desire to avoid being caught by a speed camera
*e* = my car's travelling at less than 50 mph
*f* = my not getting caught by a speed camera

Here *a* causes *b* and *b* causes *c*: we naturally deduce, plausibly enough, that *a* causes *c*. Likewise, *d* causes *e* and *e* causes *f*, so we likewise conclude that *d* causes *f*. We do not, however, want it that *a* causes *f* and that *d* causes *c*. However, if *b* = *e*, and we continue to accept transitivity, we must conclude this. We might therefore conclude that *b* ≠ *e*. Yet, if I am travelling at 30 mph, is my travelling at that speed at that time really distinguishable from *b* and *e*? Does 'my travelling at more than r mph' refer to a different event for each of the uncountably many real numbers r between 0 and 30? Surely not, even though we have uncountably many propositions <I travelled at more than r mph>, where 0 < r < 30, and perhaps also uncountably many parallel facts. So we must conclude that *b* = *e* after all, and transitivity thus fails.

Here is a similar transitivity failure that derives from the fact that a total movement can be decomposed into independent vectors of movement. Consider a device D1 where the vertical and horizontal positions of a rod perpendicular to both axes is controlled by two different levers V1 and H1. A second, identical mechanism D2, with levers V2 and H2, is connected to this, back to back, so that the two rods are joined into one single rod R situated between the V1/H1 and V2/H2 pairs of levers. Manipulating V1 thus controls the vertical component of the rod's position, which in turn controls the position of the V2 lever. The causal connection between V1 and V2 behaviour is thus clear and obvious. Ditto between H1 and H2. However, manipulating the H1 lever has absolutely no effect on V2. So it presumably follows that the behaviour of H1 is not part of the cause of the behaviour of V2. But on the Ned Hall view, it is, since it is part of its complete causal history. Observe that H1's behaviour affects the position of the rod, and likewise that V2's position was caused by the rod's being specifically where it is: both these connections are clearly causal. Given transitivity, we must, therefore, get a causal connection.

I think we have to deny causation here, for not only do we not get counterfactual *dependence*; we seem to have no *influence* of any kind—and what is causation without influence? The putative causal connection between H1 and V2 not only fails the counterfactual condition; it also fails to satisfy manipulation theories of causation, nomological theories, and just about every other theory. It is not just that horizontal positioning fails to *explain* vertical positioning. There is a clear failure of causal *production* as well: jiggle the H1 lever as much as you like, the V1/V2 parallelism remains completely unaffected. If that does not prove causal irrelevance, then what does? We seem, therefore, to have a clear counterexample to transitivity—unless, of course, we retreat to aspect-causation, and treat the horizontal and vertical vectors of the movement of the connecting rod as causally completely independent of each other.

Why did we feel in the first place that causation just has to be transitive?[1] David Lewis makes it a trivial truth, for the causal relation is just *defined* to be the ancestral of (a certain kind of) counterfactual dependence. However, this is unsatisfactory: the potential counterexamples ensure that, whatever else it is, causal transitivity is not a trivial truth. The real reason, I think, arises from the fact that most ordinary causal relations are mediate. Between any cause C and effect E, there will nearly always be underlying, intermediate mechanisms which scientific investigation will uncover. Moreover, we naturally think that this sequence of underlying, intermediate mechanisms is what actually *constitutes* the (original) causation of E by

[1] There has been remarkably little discussion of this. Cei Maslen and Ned Hall (in Collins *et al.*) present something like the argument below. Nobody, as far as I am aware, has explored the matter in any detail. There have been plenty of objections to transitivity which have been rebutted by its defenders, but the original presumption of transitivity is seldom defended explicitly.

C: ordinary causal phenomena supervene on the underlying mechanisms. After all, if there were always more to it, then the scientific analysis would turn out to be incomplete in some respect: but what additional consideration could possibly complete it? Nothing springs to mind. So unless we are allowed to presume transitivity, the science would inevitably be incomplete in this way. Thus suppose I explain why C causes E by invoking a detailed sequence of causal intermediaries $X_1$, ..., $X_n$, and show precisely how each link in the chain connects to its neighbours. This is the kind of thing that science typically tells us, and we ordinarily find this to be exactly the sort of thing we want to know about. But suppose it is then objected that all that has been shown is that C causes $X_1$, $X_i$ causes $X_{i+1}$, for $0<i<n$, and $X_n$ causes E. I need to ensure that there is no transitivity failure as well, and it is unclear what additional fact, in general, could secure this. I can check that there is no obvious funny business involved, of course, but what could guarantee the original presumption of transitivity? The most natural answer is that this is a pseudo-problem. There is nothing more that needs to be done, no additional fact that guarantees an overall causal connection over and above the causation between each link of the chain (except, perhaps, the absence of funny business that might arise in a limited set of peculiar cases). That is why, for example, we are sure that (in the car example, above) *a* causes *c* and that *d* causes *f*. Doubts here would be absurd and destructive: and this is surely because causation just is, manifestly, a transitive relation 'in the ordinary course of events', as we say, and for the reason just given.

This sounds sensible enough. The trouble is that the 'funny business' is not restricted to a few special sorts of example. 'Causal crossovers', as I shall call cases where independent causal chains pass through common intermediate events, are virtually everywhere. Thus consider this case: A says something to B who responds; in the same room at the same time, C says something to D who likewise responds. The A to B and C to D communicative exchanges are clearly causal, even though A's utterance does not cause D's response, nor does C's utterance cause B's response. However, both chains include as a common element the complex vibration of the intermediate air. The soundwaves could be decomposed into vector components, of course (the A-utterance waveform plus the C-utterance waveform, plus many other waveforms as well if it's a noisy room), but this is not especially easy to do. It is not as easy as decomposing a diagonal movement into vertical and horizontal components, for example. Nor, indeed, is there a unique decomposition, given that there is no privileged vector basis: any arbitrary waveform can be considered to be a part of the total waveform.[2] Unless our 'aspects' are seriously ethereal, causal processes will just not be 'thin' enough to avoid gross interferences.

It gets worse. We are envisaging the intermediate chain between the causal termini as lying more or less along a straight line, and this will automatically restrict what can interfere with what. But what goes on in one spatiotemporal zone is bound to have *some* effect on what goes on in any contiguous zone immediately afterwards. There will therefore be a causal chain between arbitrary C and E which passes along *any* spatiotemporal route connecting them (unless you have to travel faster than light to go that way). Even causal transmission across empty space is not obviously prohibited: after all, there is certainly a counterfactual dependence between the fact that there is a vacuum at a given zone and the fact that there is also a vacuum at a zone adjacent to it. As long as we allow absences and omissions (of which vacua form a special case) to count as causes—which is controversial, of course, but hard to avoid—it may be difficult to rule this out. Perhaps this is to overdo matters, but we need not press the point. There may be some restrictions we can introduce here, but it is clear that causal crossovers are virtually everywhere: so unless we narrow causes into pure aspects, then we must either conclude that transitivity fails in

[2] Just as an arbitrary function f(x) is a 'part' of any other arbitrary function g(x): after all, g(x) [whole] = f(x) [part] + (g(x) – f(x)) [complement], and who is to say that g(x) – f(x) is not present? A similar problem (about parts and wholes of sufficient conditions) is considered below.

general (and not just in a few odd cases), or else that virtually anything causes anything.

There are other types of example as well. Suppose we have a world that consists entirely of elementary particles which attract each other by inverse square laws (e.g., gravitational and/or electromagnetic attraction). There is a theorem (I think) which says that from the exact trajectory of just one such particle you can deduce the behaviour of all the others (given a few additional restrictions). The interactions are so all-pervasive that everything influences everything else in a very precise way. Yet even though the system is thoroughly deterministic, it seems impossible to divide it up into distinct events such that we can say, nonarbitrarily, that some of them cause others and some do not. Unless the concept of causation is to collapse into triviality, we must avoid saying that any event C which happens before E (or which is a part of event E's reverse light-cone, if we introduce a relativistic constraint) is a cause of E. Causes and effects had thus better be vastly 'thinner' than fully described trajectories of particles.

## 3 Aspect Causation and INUS Conditions

So suppose we agree to narrow causes into pure aspects: will that not solve the problem? The difficulty here is that fine-grained instantiations will not, on their own, produce anything at all (except hyperdisjunctive effects, perhaps). Many other instantiations are also going to be required if they are to generate anything of interest, and we shall have greater than usual difficulty in seeing how to avoid some other well known problems about causal irrelevance.

The problems in question are most obvious when we attempt to analyse causation in terms of necessary and sufficient conditions. This tradition, which stems from Mill (though is implicit in Hume), is not as popular as it once was, but it still deserves to be taken seriously. The most obvious candidates for being a necessary or sufficient condition for something are statements individuated up to logical equivalence. A less fine-grained entity, such as a Davidsonian event, could not be thought of as a sufficient condition *per se*: it would have to be a sufficient condition under one description, but not under another. On Davidson's view, singular causal relations, such as *c*'s causing *e*, always instantiate a law that says that occurrence of an event with property C is necessary and sufficient for the occurrence of an event with property E. As long as *c* has C and *e* has E, we have our singular causal relationship, but it is not always obvious what the properties C and E are going to be (this underlies his famous argument for anomalous monism with respect to psychophysical connections). Because the individual events *c* and *e* have coarsely grained identity criteria, they could be identified by descriptions far removed from C and E. Nevertheless, an advantage of this approach is that we can retain a version of the view that causes are necessary and sufficient for their effects. We do not need to get bogged down with well known subtleties about whether the cause has to be wholly or just partially sufficient, whether we mean (absolutely) 'necessary', or just 'necessary in the circumstances', and so forth. The causal law that instantiates the singular causal relation has a very simple, biconditional form, and these other worries can be dealt with by making a suitable distinction between a partial cause and a partial description of a cause.

This latter distinction evidently only makes sense when dealing with entities with 'secret lives' (as Helen Steward, in *The Ontology of Mind*, puts it) that can be described in many different ways—i.e., have coarsely grained identity criteria. With conditions themselves, construed propositionally, we do not get this kind of secrecy. On the contrary, what you see is what you get. If the condition is <The match was struck> then that is all of the condition. To add that the match was struck with such-and-such a force, in such-and-such an oxygenated environment, with no automatic sprinkler nearby, and so on and so forth, is not to redescribe the original condition in more detail *à la* Davidson, but to add extra, independent conditions. So understood, that the match was struck was indeed only part of the whole cause of the match's lighting: it is not a partial description of the

whole cause, for it alone was not wholly a sufficient condition for the match to light: other (genuinely independent) conditions also needed to be present.

It is with this point in mind that many have talked here of INUS (insufficient but necessary parts of unnecessary but sufficient) conditions, and the like.[3] On this view, a cause is treated as not wholly necessary and/or sufficient for its effect. Rather C causes E iff it is a nonredundant part of a sufficient condition for E (plus a time-constraint, since C must happen before E). In the above example, that the match was struck was not by itself sufficient for the match to light: the other conditions were also required. However, those other conditions on their own would not have sufficed to light the match. To that extent, the striking made a real difference, and that is what its causal role consists in. If we are to insist that causes are aspects of events, rather than events *per se*, as our earlier argument seems to make inevitable, then we shall likewise have to insist that causes cause only by virtue of their being parts of something larger.

The difficulty is in making clear the relevant notion of partiality. It may seem obvious enough: if we add extra conditions, then we get a total condition of which the original is just a part. If the total condition is sufficient, but the extra conditions on their own are not, then the original is a nonredundant part of the total sufficient condition. Unfortunately, this notion is liable to be trivially satisfied. Thus suppose that

C is an INUS condition for E

It follows that there is an additional set of factors, summarized as X, such that

{C, X} forms a (total) sufficient condition for E

and

X (on its own) is not a (total) sufficient condition for E

Let D be some (earlier) condition that genuinely obtains, but chosen at random. Then

{D, C ∨ ¬D, X} is a (total) sufficient condition for E

since it includes {C, X} (thanks to the disjunctive syllogism D, C ∨ ¬D |– C), and yet

{C ∨ ¬D, X}

is, in general, not. The addition of D makes all the difference since it blocks off the second disjunct, which ensures that D is a nonredundant part of this total sufficient condition, despite having been plucked from total obscurity. Thus D causes E, an absurdity.[4]

It may be protested that this problem is too artificial to be taken seriously. After all, it could just as easily be used to argue against the intelligibility of any sort of partial entailment. We ordinarily allow arguments to be acceptable even if they contain suppressed premises. Thus when we say that B follows from A, we often do not mean literally *that*: we mean, rather, that A, together with some suitable background assumptions that we did not bother mentioning, implies B. But then, someone objects, it follows that any proposition A enthymematically implies any proposition B: just take B ∨ ¬A as the suppressed premise. Nobody takes this objection seriously, since arbitrary B ∨ ¬A is not going to be taken by anyone to be a relevant

[3] Notably J.L. Mackie (1974)

[4] A version of this argument may be found in Kim 1971.

background assumption, even if it is known to be true. The notion of enthymematic implication surely remains more or less unscathed even if it lacks a formal definition.

I think the problem needs to be taken more seriously, however. Obviously, arbitrary disjunctions should be viewed with suspicion, but it is not easy to find a principled reason for disallowing them. What is crucial is that it is especially hard to do so if we insist that causes are fine-grained entities.

There is already a debate as to whether there can be irreducibly disjunctive facts and/or events. Disjunctive events certainly seem odd. I can (perhaps) conjoin two arbitrary events, such as the rotation and heating of a given metal ball (at time *t*). The result is presumably the mereological sum, the rotation-*cum*-heating. If we are to pursue the mereological analogy, however, then their disjunction, the rotating-or-heating, would have to be their mereological product, the common element between them (or the sum of all events that are parts of both). It is not obvious that there is any such event: the rotation and the heating are intuitively disjoint. Disjunctive facts are somewhat less odd: one may certainly cite the fact that the ball either rotated or heated at *t*, but it is nonetheless unclear that it has an independent existence. Should the ball have rotated but not heated, is the disjunctive fact independent of the fact that it rotated there and then? We may agree that

> the proposition that the ball rotated at $t \neq$ the proposition that the ball either rotated or heated at *t*

and that

> the property of heating ≠ the property of heating-or-rotating

but it is less obvious that we should accept that

> the fact that the ball rotated at $t \neq$ the fact that the ball either rotated or heated at *t*

(where the ball did not heat at *t*). If we suppose a parallel between (true) propositions and facts, then we have to accept this. But if we think of facts as more worldly items, as is not unnatural, then we may well doubt that the disjunctive fact has any reality over and above the former.

This is significant, since it goes some way towards explaining why disjunctions cannot be used to trivialize the INUS relation in the manner described above. It ceases to be true that {C ∨ ¬D, X} is insufficient for E (in the above example), because the disjunctive fact C ∨ ¬D is not really distinct from the nondisjunctive fact that C (the other disjunct is not satisfied), and C *is* sufficient for E. Thus D fails to be nonredundant: when we look at how the remainder of the total condition is actually realized, we see that it is doing *all* the work. It follows that D deserves no credit at all in the bringing about of E, which is exactly the result we want. This suggests a general, principled way of distinguishing genuine (partial) causes from the impostors: the impostors require irreducibly disjunctive collaborators that a less fine-grained identity criterion for conditions manages to exclude from the system.

It still remains to be seen whether this covers all cases. One general problem is the logical equivalence between A and (A & B) ∨ (A & ¬B). If A and B are both true, then the above argument suggests that A and A & B must stand for the same condition (the disjunction collapses into its one satisfied disjunct). Given that B is logically equivalent to (A & B) ∨ (¬A ∨ B), we can apparently show that any arbitrary two conditions A and B are in fact the same! The obvious reply is that '(A & B) ∨ (A & ¬B)' and '(A & B) ∨ (¬A ∨ B)' are not 'really' disjunctive in form, but it is unclear what saying this amounts to. They are not irreducibly disjunctive as they stand, of course, but it

is unclear what prevents a different choice of atomic formulae, for example. Moreover, we do not require our nondisjunctive facts to be saturated with *every* detail: the fact that the ball rotated at *t* might include the fact that it rotated at such-and-such a precise speed, but if we go down to full atomic details we shall find that too many other things will be included (the heating as well, most likely, since that is just an aspect of molecular movement). It is unclear how far down we can go, and there are many unsolved problems here. Nevertheless, there seems to be *some* sort of intuitive criterion at work here, even if we lack a full theoretical understanding of what underlies it; and it might be insisted that this is enough.

The trouble—and this is the crux of the matter—is that causal aspects have to be very disjunctive indeed if we are to avoid chaos, as we have already seen. The vibration of the air that is intermediate between two distinct causal chains will have an enormously complex waveform, and what is causally relevant to just one such chain will be buried amongst a mass of irrelevance. The pattern could be different in numerous ways and yet still carry the information that is essential to this one chain. The causally relevant property of the air is therefore going to be massively disjunctive—certainly so if it is to be described in exact mechanical terms (as is, presumably, required if the description is to be scientifically illuminating). It clearly also follows that, although we wish to be rid of irrelevant detail, we need many other factors to ensure that our aspects manage to produce anything at all. Each aspect on its own is so weak and disjunctive that it will not actually bring anything about except in conjunction with numerous other attendant conditions—conditions that will themselves also be highly disjunctive. So how do we prevent the impostors from entering the system? We cannot rely on an intuitive distinction between what is artificially disjunctive and what is not when everything is massively disjunctive.

One noticeable fact, which may prove helpful here, is that it is only the causal intermediaries that tend to present a problem. The original causal termini tend to be unaffected, and even where they are straightforward, as in the earlier example where A's utterance causes B's response, the intermediaries are not. We are evidently relying on intuitive criteria of macroscopic causation that we are able to apply directly and in advance of the more subtle microscopic types of causation that underlie ordinary causal processes. We have been hoping to find a single set of criteria that will do the job, but it is not obvious that this will happen. Anscombe ('Causality and Determination') famously argued that there is no single definition of causation, and the real work is done by 'thick concepts' such as *cut*, *hear*, *hit*, *pull*, and so forth. We cannot divide cuttings, hearings, hitting and pullings, and so forth, into causal and noncausal components (any more than we can separate ethically thick components into factual and evaluative components), and we can tell directly by just looking (*pace* Hume) that there is causation at work here. We know that A's utterance causes B's response in part because we know that B *heard* A's utterance. This we see pretty well directly; and once we apply the concept of hearing here, there is no room for doubt as to whether what B heard was *caused* by A's utterance: of course it was, for that is obviously part of what is meant by *hearing* something! It could be that the underlying microscopic world is as complex and as interconnected as the system of particles all subject to inverse-square laws mentioned above. It could likewise be that there is no principled way of distinguishing causal from noncausal routes through this world if we focus solely on trajectories and their vector components: the disjunctions will just overwhelm us. Yet our macrocausal judgements are unaffected; the reason is that they are underpinned by a whole raft of complex concepts that do not have a single, simple analysis. This raft is crucial, and it is possible, for all I know, for different cultures to have very different rafts here—even if they were to agree on all the microphysical laws and on exactly how the macroscopical properties depend on microscopical ones. So even if they are we agreed completely on science, we would still have very different views on what causes what.

This may sound like a rather negative conclusion (as well as deeply speculative). But counterfactual analyses are unlikely to deliver all that much, if only because we lack a general understanding of the counterfactual conditional. Nelson Goodman's ('The Passing of the Possible') analysis relies on a kind of partial entailment from antecedent to consequent; but that gives rise to all the problems raised about INUS conditions mentioned above. David Lewis's relies on an all-embracing notion of similarity; and although it is true that we have reasonably good intuitions about which of two possible situations is more similar to a third, they are subject to many variables. Lewis himself recognizes the danger. Thus 'If Nixon has pressed the button, there would have been a nuclear holocaust' is analysed as (roughly) 'In all possible worlds in which Nixon pressed the button which are most similar to the actual world, there is a nuclear holocaust'. It is then objected that worlds that have been through a nuclear holocaust are all vastly dissimilar to the actual world: worlds that have not been thus destroyed thanks to an extraordinary intervention of some kind are, all in all, much more similar. The counterfactual thus has to be judged as false, and likewise the claim that Nixon's not pressing the button caused the non-occurrence of the nuclear holocaust—which is surely absurd. Lewis's reply is:

> The thing to do is not to start by deciding, once and for all, what we think about similarity of worlds, so that we can afterwards use these decisions to test [the analysis].... Rather, we must use what we know about the truth and falsity of counterfactuals to see if we can find some sort of similarity relation—not necessarily the first one that springs to mind—that combines with [the analysis] to yield the proper truth conditions. (Lewis 1979: 43, quoted in Collins *et al.*: 6)

He goes on to develop a possible strategy to deal with this specific problem. However, it is clear that we are going to have rely strongly on antecedent intuitions about what causes what if we are likely to succeed in identifying the right kind of similarity. This suggests that a genuinely reductive analysis of causation, one where causal statements are translated accurately into statements that employ only concepts that can be grasped independently of that of causation, are unlikely to succeed. At best, we can only find useful connections between comparatively vague concepts such as causation, counterfactual dependence, similarity, relevance, artificial disjunctiveness, and so forth.

## 4 CONCLUSION

What then of transitivity, our original question? It may be difficult to get a simple answer here. The double-prevention cases, which tend to be most discussed, evidently involve problems that a simple retreat to aspects will not solve. The non-transitivity of the counterfactuals provides much of the difficulty here.[5] In the other cases, a retreat to aspects will narrow causal transmission sufficiently to prevent many difficulties, but at the price of being unable to specify just what the aspects exactly are. They are just too disjunctive to be grasped. We can retain the general presumption of transitivity, but we may need to relax the requirement, mentioned earlier, that the causation of E by C be nothing over and above the sum of the mediate causal connections between them (running from C through $X_1$, ..., $X_n$, to E). What may well sustain the fact that C causes E is that it exhibit some 'thick' causal properties that are generally recognizable, but which cannot be reduced to general properties about counterfactual dependence, INUS conditions, and so forth. What makes the process from C to E causal is indeed the fact that we have causal intermediaries, but it is also true

[5] The relation 'A is an INUS condition for B' is likewise not guaranteed to be transitive. True, if {A, X} is sufficient for B, and {B, Y} is sufficient for C, we may certainly deduce that {A, X, Y} is sufficient for C. But just because X is not sufficient for B, and Y is not sufficient for C, we may not automatically deduce that {X, Y} is not sufficient for C. Total sufficiency is obviously transitive, but nonredundancy need not be thus transmitted.

that what make intermediaries causally relevant is the fact that they form part of the larger process. Lose sight of the latter, and the integrity of the process will be lost in a sea of causal crossovers. Exactly what this leaves us with is unclear.

## REFERENCES


Collins, John, Ned Hall, and L.A. Paul (eds) 2004. *Causation and Counterfactuals*. Cambridge MA: MIT Press

Hall, Ned 2004. 'Causation and the Price of Transitivity', in Collins *et al.*: 181–203

Kim, Jaegwon 1971. 'Causes and Events: Mackie on Causation', *Journal of Philosophy* 68: 426–41

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

Mackie, J.L. 1974. *The Cement of the Universe: A Study of Causation*. Oxford: OUP

McDermott, Michael 1995. 'Redundant Causation', *British Journal for the Philosophy of Science* 46: 523–44.

Paul, L.A. 2004. 'Aspect Causation', in Collins *et al.*: 205–224