Australasian Journal of Philosophy
Vol. 79, No. 3, pp. 409–421; September 2001

LEWIS'S 'CAUSATION AS INFLUENCE'

Igal Kvart

In his 'Causation as Influence',[1] David Lewis proposed a counterfactual theory of cause which was designed to improve on his previous account.[2] Here I offer counter-examples to this new account, involving early preemption and late preemption, and a revised account, which is no longer an influence theory, that handles those counter-examples. Lewis's new account attempts to resort to aspect influence when direct counterfactual dependence as well as cause transitivity fail. Yet, I argue, both Lewis's account and the revised account fail to set apart being a cause of the occurrence of the effect and causing certain aspects of it. I present a counter-example to both accounts that brings forth this flaw.

I. Lewis's Account

On Lewis's earlier theory, *c*'s being a cause of *e* consisted in counterfactual dependence, namely, that had *c* not occurred, *e* would not have occurred, which constituted a sufficient condition for being a cause, further strengthened by cause transitivity, that is, by taking the ancestral of the counterfactual dependence relation. The major problem that faced the older theory was late preemption. According to Lewis: 'It's very clear what we want to say: one of the two potential causes did cause the effect, the other one didn't. Call the one that did the causing a *preempting* cause of the effect. Call the other one a *preempted* alternative, or backup.' (ibid., p. 182)

In the present theory, the main ingredient is Lewis's notion of influence. Consider an *alteration* of an actual event *f* to be a very fragile actual version of *f* or a very fragile non-actual alternative to *f* which is similar to *f* (ibid., p. 188). *c influences e* if not-too-distant alterations of *c* (in a substantial range, including the actual alteration) counterfactually yield alterations of *e*, at least some of which differ. (By '*g* counterfactually yields *h*' I mean here that *h* counterfactually depends on *g*.) Influence thus admits of degrees. On Lewis's position, *c* is a cause of *e* iff *c* has an influence on *e* (again, closed under cause transitivity; i.e., cause is the ancestral of the influence relation).

However, even though the above quote implies that Lewis's desideratum for the late preemption case is that the analysis yield that the preempted cause is not a cause, the quote below seems to recognise it as a cause, though much less of a cause than the preempting cause. Thus, in the famous late preemption example, where Suzy's throw of a rock (*c*) was a cause of the bottle's shattering (*e*) but Billy's throw of a rock (*b*) wasn't

[1] *The Journal of Philosophy* XCVII (2000), pp. 182–97.
[2] 'Causation', *Journal of Philosophy* 70 (1973), reprinted in his *Philosophical Papers*, Vol. 2, (Oxford University Press: 1986), and its 'Postscripts to "Causation"' there.

since it arrived just a bit too late, *c* has a lot of influence on *e*, on Lewis's view, but *b* (near enough) doesn't. Thus Lewis says:

> Now we are in a better position than before to say that Suzy's throw is much more of a cause of the bottle's shattering than Billy's. Even if the throws are so much alike that removing Suzy's throw altogether would make little difference to the shattering, it's still true that altering Suzy's throw slightly while holding Billy's fixed would make a lot of difference to the shattering, but altering Billy's throw slightly while holding Suzy's fixed would not. Take an alteration in which Suzy's rock is heavier, or she throws a little sooner, or she aims at the neck of the bottle instead of the side. The shattering changes correspondingly. Make just the same alterations to Billy's preempted throw, and the shattering is (near enough) unchanged. (ibid., p. 191).[3]

So Lewis seems to vacillate between two positions regarding whether the preempted cause should not count as a cause at all or merely as much less of a cause than the preempting cause. Even though the notion that causation admits of degrees seems right, only causes admit of degrees—non-causes do not. I thus find an outcome whereby the preempted cause does count as a cause, though much less of a cause than the preempting cause, intuitively wrong, exhibiting a treatment where one is driven to let one's theory overrule robust pre-theoretic intuitions. If indeed, as I think, in the late preemption case the preempted cause is not a cause, then I think it's at least misleading to consider the preempting cause as more of a cause than the preempted one. In cases of early preemption as well, not just of late preemption, where the time interval that makes the difference

[3] Another option that might be considered if one wishes (as one should, in my opinion) to opt for an outcome whereby the preempted cause is not a cause, is to adopt a threshold constraint (Lewis is fully aware of this option) so as to rule out as causes cause candidates (i.e., ones with some influence) with too little influence, where the threshold presumably is (at least, it seems, it should be) context-dependent. But then it doesn't seem that context yields the requisite threshold, that is, so that Billy's throw here (or in similar cases where the preempted cause has more, even much more, influence) would come out as below the threshold, and yet in other cases, in the same context, events with as much influence (or less) which intuitively count as causes, wouldn't. (See the counter-examples in section II, which illustrates how the preempted cause might have a lot of influence, which substantiates the point that the threshold option is not promising.) Setting aside redundancy cases as special cases as part of the analysis is one thing, whereas expecting context to differentiate them regarding the threshold from non-redundancy cases is another.

The above option of adopting a threshold would, moreover, bring out further the non-objectivity of the notion of cause under this account, which reflects Lewis's position in previous accounts as well, primarily through the use of similarity for the counterfactual construction which serves as the corner stone of all of Lewis's conceptions of cause, earlier ones as well the present one, since the similarity relation (which underlies the closeness relation) is taken to be context-dependent, which therefore also renders Lewis's account of counterfactuals context-dependent . (This remains so even when the properties in question are taken to be natural properties, since they need to be assigned weights or ordering.) For an approach on which the cause relation comes out objective (as do the core-cases of the counterfactual construction), see my 'Cause and Some Positive Causal Impact', *Philosophical Perspectives* 11, *Mind, Causation and World*, J. Tomberlin (ed.), 1997, pp. 401–32; my 'Causation: Counterfactual and Probabilistic Analyses', in *Counterfactuals and Causes*, J. Collins, N. Hall, L. Paul (eds.) (Cambridge, MIT Press, forthcoming); my 'Causal Relevance', in *New Studies in Exact Philosophy: Logic, Mathematics and Science* (selected contributions to the *Exact Philosophy* conference, May, 1999), Vol. II., Bryson Brown (ed.) (London, Hermes Scientific Pub. Co., 2000, pp. 59–90); and my *A Theory of Counterfactuals* (Indianapolis, Hackett, 1986), esp. chs. 2 and 9.

which potential cause ends up as the preempting cause is very small, the preempted cause has influence on the effect as well, since alterations that would result in its coming in first yield a lot of influence (see the main counter-examples below), and thus should also come out, on this position, as a cause, a robustly counter-intuitive outcome.[4] Indeed, I do agree with Lewis's position reflected in the first quote. However, the counter-examples below do not hinge on this difference that sets apart the two positions between which Lewis seems to vacillate.

There is an obvious way for Lewis to adjust the influence theory so as to have the preempted cause in the preemption cases not count as a cause if indeed it comes out as having less influence than the preempting cause, with a small price to be paid. This can be done by counting asymmetric redundant causation as a special case. That is, have the theory for cases other than asymmetric redundant causation remain as above. In cases of asymmetric redundant causation there are, say, two (there could be more) contenders for being causes (defined counterfactually in the usual way). The effect has whether-whether dependence, i.e., counterfactual dependence, on neither one of the contenders; but the one contender with more influence would now count as a cause, the other(s) would not. This is a winner-takes-all resolution of asymmetric redundant causation within the influence theory of cause. Thus, if the preempting cause comes outs, as it should, with more influence on the effect than the preempted cause, then the preempting cause comes out as a cause and the preempted cause does not, even if it has some influence. I would consider this an improvement on Lewis's account, ridding it of the counter-intuitive outcome that in all such cases both the preempted cause and the preempting cause count as causes, though of different degrees. But again, in the counter-examples below I will not assume this move, nor does this move seem to be capable of saving Lewis's new account from those counter-examples.

I now move to counter-examples to Lewis's theory of causation as influence.

II. Counter-Example about Early Preemption

Lewis's older accounts dealt neatly with early preemption. Not so, it seems, for the theory of causation as influence. Consider the following example of early preemption. A powerful sports car and a very old, fragile and slow motorised wagon participated in a race. The go signal was given. The two potential causes are the sports car's embarking on its route (*b*) and the wagon's embarking on its own route (*c*), and the effect is the cutting of the ribbon on the finish line (*e*). The wagon, while on its route, after a certain significant distance from its starting point, activated a barrier down the route of the sports car, very close to the finish line, which made it impossible for the sports car to proceed. The barrier was activated just as the sports car reached it. So the wagon cut the ribbon on the finish line while the sports car was stuck behind the barrier. Of course intuitively *c* was a cause of *e*, *b* wasn't.

[4] In addition, the gravitational pull that Billy's rock applied to Suzy's rock also had an influence, though a very slight one, on the window's shattering; and similarly so did other material objects in the area. If there is no cutoff point below which having influence does not yield being a cause, the presence of each of such objects was also a cause of the bottle's shattering, an absurd result. See note 3 above for more on the option of introducing a threshold for influence.

Furthermore, the sports car travelled at a slow speed. It could have travelled a lot faster, but the sports-car driver underestimated the wagon's maximal velocity. Had the sports car travelled a bit faster (or more than a bit), it would have passed the barrier area before the barrier closed and would have reached the finish line first. The wagon, on the other hand, was moving at the very maximal speed compatible with its constitution. Had it moved any faster, it would have broken down and fallen apart and thus wouldn't have reached the finish line at all. For simplicity, assume that each participant had to settle on a constant velocity right at the start:[5] it was not allowed to change velocity at mid-course, or else the vehicle would have been stopped. Finally, both the wagon and the sports car crossed the starting point just as the go signal was given in accordance with the rules. If one of the participants started on time but the other didn't (whether he started too early or too late), the violator would be stopped, the other would proceed. (For simplicity I include starting off too late as well, even though the above would also do for starting off too early only.) The intuitive result that *c* was a cause of *e* yet *b* wasn't remains intact.

There is a whether-whether dependence of *e* on neither *c* nor *b*. The sports car moved at a constant velocity *V*. Now consider the following (rough) velocity span for it, (*V*+, *V*+Δ*V*): velocities above *V*+Δ*V* are too distant, where Δ*V* is a substantial interval; and velocities above *V*+, which is just a bit higher than *V*, are such that the sports car would have reached the barrier before it closed shut. (Assume, for simplicity, that the sports-car driver was under strong pressure not to go below velocity *V*; had he done so, he would have given up right away and stopped his car. So alterations of *b* with velocities *under V*+ counterfactually map into the actual alteration of *c*. Thus, the subspace of alterations of *b* with velocities under *V*+ confers, by itself, influence 0 of *b* on *e* (one outcome subspace).[6] But the subspace of alterations of *b* with velocities within the first interval above (in the parentheses, which is substantial) yields a rich influence of *b* on *e*; and for *each* such velocity, that influence is pretty much as rich as that of *c* on *e* for the actual velocity V_1 of the wagon.

Regarding the wagon, alterations of *c* with velocities higher than V_1 would all map into the alteration of *e* in which the sports car cut the ribbon on the finish line, because then the wagon would have broken down very quickly (much before it would have triggered the barrier). So the alterations space of *c* with the wagon travelling faster than it did carries influence 0 of *c* on *e*. If the wagon had travelled at a speed lower than V_1, the barrier would have closed only after the sports car passed it. So the velocity alterations space in which *c* has rich influence on *e* is just V_1. On the other hand, velocities of the sports car higher than *V*+ (and lower than *V*+Δ*V*) constitute alterations of *b* that yield rich influence, regardless of the velocity of the wagon. So all in all, the non-cause *b* seems to have a lot more influence on *e* than *c*, which is a cause of *e*.

The possibility of different times of crossing starting point doesn't make a pertinent difference. To check influence here, we consider only alterations of *c* while holding *b* fixed or alterations of *b* while holding *c* fixed. With a starting time of *c* (i.e., of *c*'s crossing the starting point) that is not the actual one, *c* would be mapped into the alteration

[5] Thus, assume that both vehicles were allotted an acceleration period just prior to the official starting time. They were at liberty to select their acceleration pattern, but were required to arrive at the starting point right at the starting time, which they in fact did.

[6] The condition for influence, i.e., that some of the values of the counterfactual mapping differ is not met.

of *e* resulting from the sports car's arriving first. With a starting time of *b* that is not the actual starting time, *b* would be mapped into the actual alteration of *e*. Thus, the rich zone of influence for *c* covers just *V* with the actual staring time for *c*, and the rich zone of influence for *b* is ($V+$, $V+\Delta V$) with the actual starting time for *b*. Hence the rich influence zone for *c* is an *n*-dimensional space (with the velocity held fixed at V_1, with its actual starting time, and with the other pertinent aspects forming the other *n* coordinates). The rich influence zone for *b* is an $n+1$ dimensional space: comparable to that of *c* regarding the starting time and the other aspects but, *in addition*, with the velocity parameter varying throughout ($V+$, $V+\Delta V$), with every alteration in this interval yielding an alteration space (with the other pertinent parameters) comparable or even isomorphic to that of *c* with the velocity V_1. (The other aspects of *b* and *c* are comparable insofar as their contribution to the influence on *e* is concerned.)[7]

So all in all, *b* comes up, according to influence theory, as having more influence, and arguably a lot more influence, on *e* than *c* does. Therefore, on that theory, *b* comes out as a cause of *e* whereas *c* doesn't, or at least *b* comes out as being more of a cause of *e* than *c*, and arguably a lot more of a cause of *e* than *c*. All of these are wrong outcomes: *c*, not *b*, is a cause of *e*. The features of this example that may single it out among early preemption cases are the constitution of the wagon regarding its incapacity to sustain a higher velocity and the counterfactual features regarding different starting times (and variable velocities). But these don't seem to make a difference regarding the clear intuition that, like other early preemption cases, the preempting cause (*c* in this case) is a cause, the preempted cause (*b* in this case) is not. This is the outcome that Lewis endorsed and that was a main highlight of his earlier theories. But the influence theory doesn't seem to get early preemption right.

[7] Thus, for simplicity assume that we fix V', some velocity of the sports car in ($V+$, $V+\Delta V$), and that there are only two other one-dimensional parameters along which the alterations of *c* and *b* can vary and thereby confer distinct alterations of *e*: suppose they are the colour and the external shape of the front bumper (assume that they can be modified by the time of departure so as to yield not-too-distant alterations). The alterations spaces of *c* and *b*, then, for V_1 and V' respectively, which map into distinct alterations of *e*, are two dimensional alterations spaces. Insofar as *c* is concerned, this 2-dimensional space is the entire space of alterations of *c* that map into distinct alterations of *e*. But for the sports car there is another whole dimension—the velocity dimension, and specifically velocities within the ($V+$, $V+\Delta V$) interval. Those three parameters, we can plausibly assume, are mutually independent, and thus there are 2-dimensional spaces of alterations of *b* and *c* respectively that confer distinct alterations of *e*, which are isomorphic. But then there is another dimension, that of the velocity parameter, with lots of values in the above interval each of which yields an alteration space of *b* isomorphic to the entire alteration space of *c*. In that respect there is a clear intuitive sense in which *b* has more influence on *e* than *c* does.

In order to drive this point home further, assume that the pertinent values of the two parameters of the colour and texture of the front bumper, regarding *b* and *c*, and of the velocity, regarding *b*, are well ordered in the neighbourhoods of *b* and *c* respectively. This can be fleshed out by way of putting restrictions in terms of the competition on how many different variations the contesters are allowed to vary those parameters through, which the participants in fact adhered to, and where violation of such a restriction would yield disqualification and thus the victory of the other participant. (On top of that, this is also an obvious conceptual possibility, and Lewis's analysis is designed to be a conceptual analysis, not limited to our actual physical confines.) In that case, the number of alterations in question is finite, and, in view of the isomorphism between the pertinent alteration spaces for *b* and for *c* regarding the colour and shape of the front bumper that can be assumed to be built into the example, the extra dimension of the velocity for *b* yields more alterations for *b* over *c* in a precise finite sense.

A last comment regarding this case and cause transitivity: Lewis has held out in favour of transitivity of cause against what seem to be compelling counter-examples.[8] There was indeed a strong motivation for doing so on Lewis's previous theories, since cause transitivity was necessary to offset the lack of counterfactual transitivity in the case of early preemption. However, we now see that influence theory does not handle early preemption adequately. Hence there does not seem to be anymore as good a reason for Lewis to defend cause transitivity: early preemption under influence theory doesn't come out right regardless of whether or not cause transitivity holds.

III. Counter-Example about Late Preemption

A small modification can lead to a similar result, but now in the standard late preemption case. Take the very same example above, but *without* the barrier. The driver of the sports car, being overconfident (in a way reminiscent of the hare that raced against the tortoise), exaggerated the considerable advantage he had, became less careful and took his time. As a result, he arrived at the finishing line a tad after the wagon.

Again, there is no whether-whether dependence of e on either c or b. Regarding the sports car, consider again its velocity V and alterations with velocities in (V+, V+ΔV). Again, velocities above V+ΔV are too distant, where ΔV is a long interval; and for velocities above V+, the sports car would reach the finish line before the wagon. Again, alterations of b with velocities *under* V (that are not too distant) counterfactually map into the actual alteration of c; so the subspace of alterations of b that fall in the velocity zone of V and below confers influence 0 of b on e. But again, the subspace of alterations of b with velocities within (V+, V+ΔV) (which is long) confers rich influence of b on e; and for *each* such velocity, that influence is pretty much as rich as that of c on e for the actual velocity V_1 of the wagon.

Alterations of c with velocities higher than V_1 would map into the alteration of e in which the sports car cuts the ribbon on the finish line, because then the wagon would have broken down. So the alterations space of the wagon travelling faster than it did carries influence 0 of c on e. Alterations of c in which the wagon travels at a speed lower than V_1 all map into the alteration of e that would result with the sports car reaching the finishing line first. So in the alterations space in which c has a rich influence on e the velocity of the wagon is fixed at V_1.

Other aspects of b and c are comparable insofar as their contribution to the influence on e is concerned. Hence, again, the rich influence zone for c is an n-dimensional space (with the velocity held fixed at V_1, with the actual starting time, and with the other aspects forming the other n coordinates). The rich influence zone for b is an n+1 dimensional space: the analogous one to that for c regarding the starting time and the other aspects but, in addition, with the velocity parameter varying throughout (V+, V+ΔV). So all in all, b comes up, according to influence theory, as having more influence on e than c does,

[8] In particular, Ned Hall's and Hartry Field's. For my counter-example, see 'Cause and Some Positive Causal Impact', section 5, and 'A Counterfactuals Analysis of Cause', *Synthese* (forthcoming), section 1. For the diagonalisation argument against transitivity of causal relevance and thus of cause, see my 'Transitivity and Preemption of Causal Impact', *Philosophical Studies* 64 (1991), pp. 125–60, section 5.

arguably a lot more. Thus *b* comes out a cause of *e* and *c* does not, or at least *b* comes out as being more of a cause of *e* than *c*, and arguably a lot more. All of these are, again, wrong outcomes: *c*, not *b*, is a cause of *e*. And again, the only features of this example that single it out *vis-à-vis* the standard late preemption case are as in the above example regarding the actual early preemption case. They do not make a difference regarding the clear intuition that, like other late preemption cases, the preempting cause, i.e., *c* in this case, is a cause, the preempted cause, i.e., *b* here, isn't, or at the very least that the preempting cause is more of a cause than the preempted one. This is the outcome that Lewis endorsed and was, it seems, of major importance in motivating the influence theory. But this new influence theory doesn't seem to get late preemption right either.

Hence, I conclude, influence theory gets neither late preemption nor early preemption right. It doesn't handle right late preemption, which was a main motivation for abandoning the earlier account; and it doesn't handle right early preemption, which the earlier account handled adequately. Hence the new influence theory is, in terms of handling redundancy cases, no advance over the previous one. Although richer in various respects and interesting, it does worse, in terms of extensional adequacy, than its predecessor.[9]

IV. A Generalised Topological Framework

As I see it, one can fit Lewis's influence theory of cause in a broader perspective of topological mappings as follows. For preemption cases, where whether-whether dependence fails, the general idea, within which Lewis's specific proposal falls, is to focus on counterfactual dependence between the *neighbourhoods* of the cause and the effect, and to aim at imposing the condition that, in such a case, *c* is a cause of *e* iff such a dependence of the right sort holds. That is, the idea is to consider not just the counterfactual function for *c* and *e* alone, but to take into account its behaviour in their neighbourhoods. This is a topological approach, which looks for cues regarding functional behaviour at a particular point through the behaviour of the function in its neighbourhood, in analogy, say, to a function that is undefined at a point through its explicit definition but whose behaviour at

[9] If intermediate events are allowed to play a role in the determination of influence, influence theory would be more, possibly a lot more, complicated than the account Lewis provided. Call the account of influence in terms of counterfactual dependence between alterations of the effect candidate and alterations of the cause candidate, as Lewis presented it, *direct* influence. If intermediate events are to play a role, two problems arise: first, for a chain (of, say, 3 elements) *c*, *d* and *e*, what is the influence of *c* on *e* conferred by the chain given the direct influence of *c* on *d* and of *d* on *e*? Second, what is the influence of *c* on *e* (call it influence *simpliciter*) given two or more such chains? Lewis does not offer answers to these problems (though his conclusion that influence is not transitive yields certain constraints). In his discussion of the example of late preemption as well as the examples he uses to argue that influence is not transitive he proceeds as if only direct influence is involved. I followed suit and so did I in the examples above. But if influence *simpliciter* is not direct influence, then Lewis's discussion of these examples doesn't establish the conclusions he reaches regarding them since there is, in his paper, no account of influence *simpliciter* (if other than direct influence). If this is indeed the case, Lewis's theory of causation as influence is, as presented in this paper, inapplicable by and large to particular examples (i.e., so long as there are intermediate events), and thus is by-and-large not testable and has no predictive power. This paper, therefore, is best considered, it seems, as a theory of direct influence, since there is no theory of indirect influence in this paper.

the neighbourhood of that point invites a natural stipulative definition (which yields, say, continuity). Lewis's notion of an *alternate* plays the role of specifying the points of which the neighbourhoods consist: the neighbourhood of *c* consists of a fragile actual version of *c* as well as fragile non-actual versions of *c* that are similar to *c*; and similarly for *e*. Thus, as in his approach to counterfactuals, similarity plays the topological role of ordering that invites the notion of a neighbourhood (compare Lewis's notion of a sphere in his characterisation of truth conditions for counterfactuals). Let a ***counterfactual mapping*** *between two neighbourhoods* of *c* and *e* be a function *f* that maps points at the neighbourhood of *c* to points at the neighbourhood of *e* such that *e′* (which is in the neighbourhood of *e*) is the value of *c′* (which is in the neighbourhood of *c*) just in case *c′>e′*. (All the points in the neighbourhoods of *c* (and *e*) are non-actual, except for a fragile actual version of *c* (and *e*), which are in the neighbourhood.)

Given this framework, the question is whether there are natural constraints on counterfactual mappings that capture a (genuine) cause among the cause candidates in the preemption cases. Call such a mapping ***degenerate*** just in case it has only one value for all its arguments (where it is defined, i.e., for those arguments for which the function has a value). Of course, the mapping can be richer or poorer depending on whether the range of its values has more or less members. Consider a counterfactual mapping ***substantive***, after Lewis, in case it assigns values to a substantial range of arguments within the neighbourhood of *c*. Lewis's account of influence can be viewed as offering specific constraints on counterfactual mappings that are designed to capture a genuine cause in the case of preemption (in particular, late preemption). It can be expressed as follows (for redundant causation cases):

(**TI**) *c* is a cause of *e* just in case there is a substantive counterfactual mapping of the neighbourhood of *c* into the neighbourhood of *e* that is not degenerate.

(TI – for topological influence.)

V. Cause as Influence Preservation through Sustained Reducibility

The counter-examples above point to what seems to be a serious deficiency in the influence theory. The topological perspective laid out in the last section, within which Lewis's influence theory can be couched as a special case, i.e., as (TI), makes it possible to formulate another conception of influence, different from Lewis's, but within this general topological framework, that is capable of handling the above counter-examples. This conception, presented here, may be abstract and hard on the reader. But its application below to the above counter-examples will help clarify it and bring forth its intuitive motivation.

I will now explain this conception.

Among the various parameters involving *c*, which are *c*'s features or aspects, some are **lively** and some are **dormant** (within a given neighbourhood). A *parameter is lively in an interval*[10] (*vis-à-vis* a neighbourhood of *c*) if different alterations in the neighbourhood that differ just in their values for this parameter that are in that interval yield different

[10] I assume here and below that the parameters involved are one-dimensional parameters whose values can be considered on a linear scale.

values (in the neighbourhood of *e*) for the counterfactual mapping. A *parameter is lively in a neighbourhood (or a subset of it)* if different points in the neighbourhood (or its subset) have values of the parameter such that different points in the neighbourhood that differ from each other only in those values for the parameter in question are mapped into different points in the neighbourhood of *e*. Otherwise the parameter is *dormant*. A *neighbourhood* (or a sub-neighbourhood—see below) *is lively* if some parameter is lively in it. For our purposes, we need not bother about dormant parameters, so we can ignore them and just focus on lively ones.

Thus, return to our simple-minded version of the wagon vs. the sports car case. Assume (here and below) that the only parameters of the vehicles (other than the velocity) that affect aspects of the outcome *e* are the colour and sharpness of the front bumper. These parameters are lively (in the neighbourhood of *c*) since some modifications in their values (which are of points in the neighbourhood) yield different points in the neighbourhood of *e* through the counterfactual mapping. But the parameter of, e.g., the colour of the wheel, is dormant, since different points in the neighbourhood of *c* that differ *just* in the colour of the wheel of the wagon don't yield different values of the counterfactual mapping.

A candidate for **sub-neighbourhood** of a given neighbourhood of *c* is a subset of the original neighbourhood that includes the actual version of *c*. It must also be *non-degenerate*, i.e., it must not consist of just a single point. A candidate for sub-neighbourhood is **connected** just in case, if any two alterations of *c* that are in the sub-neighbourhood differ only in their values for a certain parameter, then any alteration of *c* that agrees with these two alterations regarding the parameters on which they don't differ but, regarding that parameter, has a value between the values of this parameter for these two alterations, is also in the sub-neighbourhood. A candidate for sub-neighbourhood that is connected is a *sub-neighbourhood*.

The main idea behind the sustainably reducible influence account is that, in redundancy cases, when *b* and *c* are the cause candidates, *b* is not a cause of *e* (and thus is the preempted cause) if, even though *b* has influence on *e*, there is a lively sub-neighbourhood of *b* in which *b* has no influence on *e*. That condition holds when it is not the case that *the influence of b on e is kept as non*-0 *through every shrinking of the neighbourhood of b* (into a lively sub-neighbourhood of *b*). However, for *c*, in such a case, to be a cause of *e*, *c*'s influence on *e* must be *sustainable* (i.e., kept as non-0) by any shrinking, i.e., downsizing, of the neighbourhood of *c* into a lively sub-neighbourhood of itself.[11] Thus, the proposed account is:

(SRI) In redundant causation cases, *c* is a cause of *e*, but *b* is not, just in case there is a lively sub-neighbourhood of *b*, but not of *c*, that confers influence 0 on *e*.

[11] Note that in order to verify that the influence of *b* on *e* is not sustainably reducible, one needs to select a lively sub-neighbourhood. To ensure that the sub-neighbourhood is lively, it's enough to find out two values of a certain parameter in an interval that is lively *in the neighbourhood in question*. That is, that there be two points in the neighbourhood in question that have different values for this parameter, have the same value for all other parameters, and are counterfactually mapped into different values in the neighbourhood of *c*. But although the sub-neighbourhood in question must include the interval of the parameter in question in order to be lively, it need not include the two points of the neighbourhood in virtue of which this parameter is lively in that interval.

Call this account a **sustainably reducible influence account of cause** (thus (SRI)). Put differently:

(SRI′) In redundant causation cases, *c* is a cause of *e*, but *b* is not, just in case there is a lively sub-neighbourhood of *b*, but not of *c*, in which the counterfactual mapping is degenerate.

We can see now that this new account, the sustainably reducible influence account of cause, handles the above counter-examples to the influence theory of early and late preemptions adequately. The main idea of the sustainably reducible influence account, applied to those examples, is that *b* (the sports car's embarking on its route) is not a cause of *e* (the cutting of the ribbon on the finishing line) since *b*'s influence on *e* is not sustainable under a shrinkage, i.e., down-sizing, of the neighbourhood of *b* into certain lively sub-neighbourhoods of *b*. That is, *b* indeed did have influence on *e*. But that influence was restricted, insofar as the velocity parameter is concerned, to the interval (V+, V+ΔV), and this interval does not include the actual velocity V of the sports car. The remaining interval of the velocity parameter in the neighbourhood of the actual fragile version of *b* is dormant, and the actual fragile version of *b* that is in the neighbourhood is in this remaining interval, and that remaining interval is connected. The velocity parameter is indeed a lively parameter of *b* in the neighbourhood. But in view of the above observation, there is a lively sub-neighbourhood of *b* that has influence 0 on *e*. Points in this sub-neighbourhood have values for the parameters of the colour and sharpness of the front bumper of the sports car that range through the values of these parameters that any point in the neighbourhood of *b* has (and include the actual values of those parameters), and therefore those parameters are lively in the sub-neighbourhood. But all the points in this sub-neighbourhood have values for the velocity parameter that are in the interval. (V',V], i.e., an interval that includes the actual velocity V and is a closed interval on the right, for any V' that is lower than V but is still a velocity value for some points in the neighbourhood. In such a sub-neighbourhood, the influence of *b* on *e* is 0, since for all the points in such a sub-neighbourhood (which are alterations of *b*), the wagon would cut the ribbon on the finishing line since the velocity of the sports car was not high enough, and therefore all these points are mapped into one and the same actual alteration of *e*.

This is not the case, however, regarding *c* (the wagon's embarking on its route). For any choice of lively intervals for the parameters of *c*, the influence of *c* on *e* in a sub-neighbourhood that contains them, and that is therefore lively, is not 0. Thus, take any intervals of the colour and sharpness parameters of the front bumper of the wagon within a neighbourhood of *c*. Since a sub-neighbourhood must include an actual version of *c*, any such lively sub-neighbourhood has a non-0 influence on *e*. This is since some alterations of *c* with the actual velocity V_1 of the wagon but with different values of the colour and/or sharpness of the front bumper are in such a sub-neighbourhood, and they are mapped into different alterations of *e*. By contrast, a subset of *c*'s neighbourhood may have influence 0 on *e* if the interval of the velocity parameter that is in it does not include the actual velocity V_1 of the wagon; but then it does not include the actual version of *c* in the neighbourhood, and thus does not qualify as a sub-neighbourhood.

The above applies equally to the early and late preemption examples discussed in sections II and III. Influence theory, as an account of cause, fails in the above counter-

examples; but the sustainably reducible influence account above does handle them adequately. Consequently, since early preemption is handled adequately by the sustainably reducible influence theory without recourse to cause transitivity, the claim that cause is transitive which, I have argued, does not hold anyhow, is not needed for this account of cause. The case of symmetric redundant causation seems to yield sustainably reducible influence for the various contenders of being causes, and thus they all come out as causes, the right outcome.

Whether-whether dependence, i.e., counterfactual dependence, which is a form of influence that doesn't hold in redundancy cases, still remains a sufficient condition for being a cause. Yet one must not require whether-whether dependence for all cases other than redundant causation cases. Thus, in some non-redundant causation cases, influence of the non-whether-whether dependence is required for a counterfactual account, in particular in cases of contributory causes. Consider the following example:[12]

> Two individuals, #1 and #2, attacked and overcame a third, the victim, on a river bank, and then dragged the victim into the water. #1 had enough strength to submerge the victim alone, even though in fact #1 exercised only a small portion of that strength, way below what was needed for #1 alone to submerge the victim. Thus, the contribution of #2 in fact made a real difference. However, suppose that the disposition of #1 was such that, if necessary, in particular, had #1 been acting alone, he would have exercised the requisite force to submerge the victim. Still further, suppose that #2 didn't exercise the requisite strength to submerge the victim alone, nor would he have had he acted alone. Consider e to be the drowning of the victim, and c_i to be attacker #i's endeavoring to keep the victim's head submerged.

Intuitively, it seems, each of the c_is was a cause of e. In particular, c_2 was a cause of e. Yet in this case, the counterfactual $\sim c_2 > \sim e$ (namely, if attacker #2 had not endeavoured to keep the victim's head submerged, the victim wouldn't have drowned) is false.

e has counterfactual dependence on c_1 but not on c_2; yet both c_1 and c_2 are causes of e. This is therefore not a case of redundant causation in which counterfactual dependence as a condition for being a cause fails. Yet c_2 does have influence on e: appropriate counterfactual modifications in various features of c_2 yield appropriate modifications in features of e. Thus, influence on e, though not of the whether-whether sort, is required for a counterfactual analysis of cause beyond cases of redundant causation.

VI. The Flaw of Aspect Over-Sensitivity

However, even though a sustainably reducible influence account such as the one suggested above seems to be capable of overcoming the problem posed by the counter-examples in sections II and III, there is another feature of Lewis's theory that is not saved by the above shift to the new account. In a nutshell, since Lewis's conception of events favours a relatively broad event individuation, there is influence, and sustainably reducible

[12] I have presented this example in my 'A Counterfactual Theory of Cause' as a counter-example against Lewis's previous theory of cause.

influence, even when the influence is directed only at *aspects* of the effect but not at all at its very occurrence.[13] When this is the case for the backup cause, and its influence exceeds that of the real cause, influence theory as well as the sustainably reducible influence account yield the wrong outcome.

To build a concrete example along such lines, consider a variation of the standard Billy and Suzy case of late preemption, in which they both threw rocks at a window which then shattered. Assume that in front of the window there was a barrier that was highly sensitive to impact. If an impact was made on the barrier, the barrier would activate a very small explosive device that would shatter the window. Once an impact was made on the barrier at some point in time prior to t, it makes no difference whether another impact, later than the first, was or was not made on the barrier: the barrier would activate the explosive device exactly at t, regardless of when (prior to t) the impact was made. Further, the barrier was not sensitive to different *ways* of being impacted; and it was capable only of activating or not activating the explosive device.

In fact, the impacting of the barrier by Suzy's rock at t' occurred (call this event $\boldsymbol{a}$), where t' was well prior to t. But Billy's throw also hit the barrier immediately after Suzy's rock, and thus also well before t. Thus, the direct influence of c (Suzy's throw) on e (the shattering of the window) was 0: all the alterations of c map into just the actual alteration of e. Hence the direct influence of c on e was 0.[14]

Now consider $\boldsymbol{b}$ (Billy's throw). Suppose that various tiny devices had been distributed around the window, say, numerous elastic and adjustable strips glued over various portions of the window. Those strips had sensors that were highly sensitive to the manner in which Billy threw his rock (though not at all to the manner in which Suzy threw hers). Further, different readings of those sensors would be translated into different distributions of the window's cohesiveness (the connection of some parts of the window to others could

[13] For a discussion of aspects, see my 'Cause: Time and Manner' (forthcoming). See also Daniel Hausman, *Causal Asymmetries* (Cambridge University Press, 1998), ch. 2, sec. 2.3; and L.A. Paul, 'Causal Aspects', *The Journal of Philosophy* XCVII (2000), pp. 235–56.

[14] To further tighten the point that the influence of c on e was minimal, consider the *indirect* influence of c on e, namely, via a, through cause transitivity. The influence of a on e was minimal: all the alterations of a map into just the actual alteration of e (the shattering of the window). This should be enough to yield that the influence of c on e via a was minimal (and so should come out the sustainably reducible influence, *mutatis mutandis*).

If further tightening of this conclusion is desired, though it doesn't seem necessary, then boost the example in the following way so that the influence of c on a will be minimal and thus the indirect influence of c on e will be *at most* minimal. Thus, assume that Suzy had to throw her stone through a filter that lets the stone through only if it has certain precise specifications involving the manner of Suzy's throw, which had been determined prior to the throw. In fact, Suzy's throw met those specifications, and the filter let Suzy's stone through, though it wouldn't have otherwise. The point of introducing the filter in this example (analogously to the resort to the barrier) is to have the effect that different alterations of Suzy's throw map into only two alterations of a, namely the actual alteration of the barrier's being impacted at t or the alteration with its non-occurrence (but still with the impact of Billy's rock), which yields minimal influence of c on a, namely, influence yielded by a counterfactual mapping that has only two values—only two distinct alterations of a. Further, in addition to the direct influence of c on a's being minimal, so is the influence on a of the filter's letting Suzy's rock through—again, it is minimal.

These wrinkles on the example are designed to maximise the contrast of the influence of b on c vs. that of c. One need not quibble over the details, however, so long as it is appreciated that b's influence is very considerable and that the example can be tailored so that c's influence is appreciably less. Then the point of the example carries over even if the influence of c on e is not recognised as minimal (i.e., yielding a mapping with only two distinct alterations of e as values).

be strengthened or weakened by adjusting the many strips in numerous ways), even though Billy's throw hit after Suzy's. Different manners of Billy's throw would thus yield different readings of the sensors, which in turn would yield different distributions of elasticity of the various strips, which in turn would result in different *manners* of the shattering of the window. Thus, many alterations of *b* map into different alterations of *e*. *b* therefore has considerable influence on *e*, but *c* has 0 (or minimal, or little) influence on *e*. Therefore, on the influence account, *b* has considerable influence on *e*, *c* doesn't, and thus *b* has a lot more influence on *e* than *c* does. Therefore, *b* comes out as a cause of *e* and *c* doesn't, which is the wrong outcome. In fact, *c* (Suzy's throw) was a cause of *e* (the window's shattering), and *b* (Billy's throw) wasn't.

Nor, it seems, is the sustainably reducible influence account immune from this example: there is no sub-neighbourhood of the neighbourhood of *b* that has influence 0 on *e*.

Therefore *b* has sustainably reducible influence on *e*, and more such influence than *c* does, and so, on the sustainably reducible influence account as well, *b* comes out as a cause of *e* and *c* doesn't, or at least *b* comes out as more of a cause of *e* than *c*, both of which are the wrong outcomes.

This type of example brings out the phenomenon of *exclusive* causes of manner,[15] i.e., cases where one event is not a cause of the latter even though it is a cause of aspects of the latter. Such a distinction is analysable on the probabilistic account of cause that I have offered elsewhere.[16] Unless a comparable distinction can be made on Lewis's conception of events in a deterministic world (since Lewis's influence theory is designed for a deterministic world) in a way that would make it possible to exclude exclusive causes of aspects (i.e., causes of aspects only), both the influence theory as well as the sustainably reducible influence account are vulnerable to the above counter-example.[17]

Hebrew University, Jerusalem Received: June 2000

[15] See my 'Causes: Manner and Time'.

[16] The account is designed for an indeterministic world. See my 'Causal Relevance'; my 'Causation: Counterfactual and Probabilistic Analyses'; and especially my 'Causes: Manner and Time'.

[17] I thank Alex Byrne and Michael Fara for comments on the content of early drafts of the first part of this paper. I also thank Carl Posy for comments on the latter part.