Counterfactual Causation and the Problem of Indiscretion

Joseph Long

On some accounts, causation is a relation between discrete entities such as events, facts, or states of affairs[1]; other accounts take causation to consist in processes such as the flow of energy or momentum, physical processes more generally, or the transference of tropes.[2] (Call the former 'discrete causation', the latter 'process causation'.) Any adequate analysis of causation must preclude one of these competing types of accounts. In "Causation" (1999 [1973]), David Lewis offers a counterfactual analysis of causation in terms of propositions whose truth conditions consist in the occurrence of events related via causal dependency.[3] I show in this paper that Lewis's analysis is too weak for it is consistent with both discrete and process causation.

An overview will prove useful. In the first section, I summarize Lewis's counterfactual account. In the second section, I show that Lewis's account entails that if an object is at rest for any duration, then its initial rest state causes all subsequent rest states. In the third section, I show that Lewis's account therefore entails that mere continuation implies causation, and thus Lewis's account fails to distinguish between discrete and process causation. I shall assume time is continuous (*i.e.,* time does not consist in finite fundamental units), and I make no assumption regarding whether discrete or process causation is correct, although my analysis favors the latter.

Causation, Counterfactual Style

Assuming determinism and setting aside INUS-conditions, David Lewis (1999 [1973]) presents a counterfactual analysis of causation in terms of particular cases and the ordinary conception of events.[4] For Lewis, a counterfactual is a relation between propositions such that

(1) $C \; \Box\!\!\rightarrow E$

if C were true, then E would be true. Now, according to Lewis, (1) is non-vacuously true iff E is true at all the closest worlds in which C is true. The notion of closeness is cashed out in terms of comparative similarity: For any two worlds, w_a and w_1, world w_1 is comparatively similar to w_a iff there exists a third world, w_2, such that w_1 is more similar to w_a than w_2 is to w_a. Therefore, (1) is (non-vacuously) true iff there is at least one world w such that C and E are both true at w, and w is more similar to the actual world than any other world at which C is true, but E is false.

Going further, suppose that there are two equal-sized families of propositions, C_1, C_2,... and E_1, E_2,... and that within each family of propositions no two propositions are compossible. If the following holds for all the members of the two families,

(2) $(C_1 \;\Box\!\!\rightarrow E_1)$ & $(C_2 \;\Box\!\!\rightarrow E_2)$ & $(C_3 \;\Box\!\!\rightarrow E_3)$ &...

then the family of *E*s depends counterfactually on the family of *C*s. Thus, as it is said, if C_1 were true, then E_1 would be true; if C_2 were true, then E_2 would be true; if C_3...; and so on. For simplicity, since the propositions within the families are not compossible, we can say that the family E, $\sim E$ depends counterfactually on the family C, $\sim C$.

The next step on the way to Lewis's account of causation is to pair propositions with events, which is done in terms of possible worlds. An event c corresponds to a proposition C iff C is true at all and only those worlds at which c occurs.[5] Given the possible worlds pairing of events and propositions, if *(a)* a family of compossible propositions, C_1, C_2,... corresponds to a family of compossible events, c_1, c_2,..., *(b)* a family of propositions, E_1, E_2,... (equal in size to the family of *C*s) corresponds to a family of events, e_1, e_2,..., and *(c)* the statement (2) characterizes the relation between the *C*-members and the *E*-members, then the family e_1, e_2,..., says Lewis, depends causally on the family c_1, c_2,.... Or, as it is said, whether e_1 occurs causally depends upon whether c_1 occurs; whether e_2 occurs causally depends upon whether c_2 occurs; whether e_3...; and so on. Or, more simply, the family *e*, ~*e* causally depends upon the family *c*, ~*c*.

Now, turning to single events, since event-family members are not compossible with each other, then the causal dependence depends upon two counterfactuals:

(3) $C \mathbin{\Box\!\!\rightarrow} E$

(4) $\sim C \mathbin{\Box\!\!\rightarrow} \sim E$

Thus, for a single event e_1, whether or not e_1 occurs causally depends upon whether or not c_1 occurs. Of causal dependence between single events, Lewis now has in hand an analysis in terms of counterfactually related corresponding propositions, where the counterfactual relation is cashed out in terms of truth-conditions across possible worlds.

Finally, according Lewis's account, given any two actual events *c* and *e*, if *e* would not have occur had *c* not occurred (*i.e., e* is causally dependent upon *c*), then *c* is the cause of *e*.[6] Given that causation is a transitive relation, we can say that if *c* caused *d*, and *d* caused *e*, then *c* caused *e*. Lewis calls relations like that running from *c* through *d* to *e* 'causal chains'. Thus, for any two events *c* and *e*, a causal chain runs from *c* to *e* iff *e* is causally dependant upon *c*. Thus, Lewis concludes his analysis, *c* causes *e* iff there is a causal chain leading from *c* to *e*.

Causation, Continuation, and a not so Discrete Problem

From Lewis's analysis of causation, one can extract the following three claims:

(5) For any two events *c* and *e*, *c* causes *e* iff there is a casual chain from *c* to *e*.[7]

(6) For any two events *c* and *e*, there exists a causal chain leading from *c* to *e* iff there is a finite sequence of actual particular events such that *e* causally depends on *c*.[8]

(7) For any two events *c* and *e*, *e* causally depends on *c* iff the family of corresponding propositions *E*, ~*E* counterfactually depends on the family *C*, ~*C*.[9]

From (5), (6), and (7), we get the following analysis of causation:

(CC) For any two events *c* and *e*, *c* causes *e* iff the family of corresponding propositions *E*, ~*E* counterfactually depends on the family *C*, ~*C*.

That is to say, one event causes another just in case both events occur, and had the former not occurred, the latter would not have occurred. The following two examples show, however, that (CC) is inadequate.

Suppose *(i)* that there are three neutrally charged particles, α, β, and γ, and *(ii)* that relative to some reference frame α and β are at rest and γ is traveling at 10 m/s. Suppose, furthermore, that two events occur: event *a* is such that at time t_1 γ passes by α at a distance of 2 m and with a linear trajectory, $f(x) = x$; event *b* is such that at time t_2 γ passes by β at a distance of 2 m and with a linear trajectory, $f(x) = x$. Finally, suppose that no other wave or particle acts on γ during the spacetime interval between *a* and *b*. Assuming Newton's first law of motion, if *a* had not occurred, then *b* would not have occurred. That is, had γ not passed by α at such and such distance and trajectory, then it would not have passed by β with the same distance and trajectory. But by (CC), if *(a) a* and *b* both occur, and *(b)* had *a* not occurred, *b* would not have occurred, then *a* caused *b*. Thus, on Lewis's account, γ's passing by α as it did *caused* its passing by β as it did. This is odd, but things get odder still. Take the next example.

Using Galilean transformation rules, the events just described can be re-described in an infinite number of ways. Relativity theory claims there is no fact of the matter about which description is correct. So, consider the following re-description: *(i)* there are three neutrally charged particles, α, β, and γ, and *(ii)* relative to some reference frame, γ is at rest and particles α and β are traveling at 10 m/s. Two events occur: event *a** is such that at time t_1 α passes by γ at a distance of 2 m and with a trajectory, $f(x) = x\text{-}(2\times2^{1/2})$; event *b** is such that at time t_2 γ passes by β at a distance of 2 m and with a trajectory, $f(x) = x\text{-}(2\times2^{1/2})$. No other wave or particle acts on γ during the interval between *a** and *b**. Assuming Newton's first law of motion, if *a** had not occurred, then *b** would not have occurred. That is, had α not passed by γ at such and such distance and trajectory, then β would not have passed by γ with at its distance and with its trajectory. But by (CC), if *(a) a** and *b** both occur, and *(b)* had *a** not occurred, *b** would not have occurred, then *a** caused *b**. Thus, on Lewis's account, α's passing by γ as it did *caused* β's passing by γ as it did. Now, this is odder than the last scenario, but things get odder still.

Suppose particles α and β are removed from the last scenario: there is a single neutrally charged particle, γ, that it as rest relative to some reference frame at times t_1 and t_2. At no time between t_1 and t_2 does another wave or particle act on γ. Newton's first law of motion, for any two times t_a and t_b where $t_1 \leq t_a < t_b \leq t_2$,[10] γ is at rest at t_a and t_b, and had γ not been at rest at t_a, then it would not have been at rest at t_b. But then by (CC), γ's being at rest at t_a caused γ's being at rest at t_b! Something has gone wrong. The remedy, I believe, is for the counterfactualist to forego discrete causation, but I shall not argue for that here. Next, though, I show that Lewis's counterfactual account fails to distinguish between discrete and process causation.

A not-so Discrete Problem

Call the event that obtains when an object is at rest at a time a "stationary event." Use the phrase 'stationary relation' to refer to the relation that obtains between two stationary events such that *(i)* the objects in the stationary events are identical, and *(ii)* there is no time between the two stationary events such that the object is not stationary. Thus, in the last scenario described, γ's being at rest at t_a is a stationary event; and since *(a)* γ is at rest at t_a and at t_b, and *(b)* there is no time between t_a and t_b at which γ is not at rest, then γ's being at rest at t_a and γ's being at rest at t_b stand in a stationary relation to each other. Following convention, if there is a continuity between two events, say that there is a process that extends from one event to the other.[11]

Give time's continuity, it now becomes clear that (CC) fails to distinguish between process causation and discrete causation. Notice, first, that given Newton's first law and (CC), it is the case that, for any stationary relation, had the first stationary event not occurred, then the second would not have occurred; therefore, all stationary relations are causal relations. Second, if time is continuous, then, for any stationary relation, that relation is continuous; that is, the earlier stationary event is continuous with the later stationary event. But by definition, if an earlier stationary event is continuous with the later stationary event, then there is a process that runs from the earlier to the later event. Therefore, on (CC), for any stationary relation, that relation is a continuous causal process, showing that *(CC) does not distinguish between discrete and process causation for stationary events related stationarily*, as it were.

Finally, returning to Galileo and Einstein, whether an object is at rest or in motion is relative to a reference frame, and there is no fact of the matter about which reference frame is correct. Applying these principles to stationary relations, there is therefore no fact of the matter about whether the object in the stationary-event relata is at rest. That is, for any stationary relation, the event-relata can be re-described such that the relevant object is in motion. But if events can be re-described as such, then there is no fact of the matter about whether stationarily related events are in fact stationary, in which case the only fact of the matter is that one event is continuous with the other. Therefore, *if (CC) does not distinguish between discrete and process causation for stationary events related stationarily, then (CC) does not distinguish between discrete and process causation for events whose objects are in constant motion, either.* In short, on Lewis's analysis, mere continuation implies causation. I therefore conclude that Lewis's counterfactual account of causation does not distinguish between discrete and process causation.

Conclusion

I have argued that David Lewis's counterfactual analysis of causation does not distinguish between discrete and process causation and is therefore too weak. After briefly describing Lewis's analysis in the first section, I show in the second section that Lewis's analysis allows for the oddity that if an object is at rest for any duration, then that object's being at rest at an earlier time in the duration causes the object's being at rest at any later time in th duration. In the third section, I show that Lewis's analysis entails that continuation implies causation, and I conclude that his analysis therefore fails to distinguish between process and discrete causation. I have not argued that process causation should be embraced over discrete causation; if, however, one finds Lewis's counterfactual account compelling but the continuation-implies-causation implication appalling, I might suggest he or she seriously consider the process accounts of causation.

References

Armstrong, D. M. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.

Bennett, Jonathan. 1988. *Events and their Names*. Indianapolis: Hackett Publishers.

Davidson, Donald. 1980. Causal relations. In *Essays on Actions and Events*. Oxford: Clarendon Press.

Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Fair, David. 1979. Causation and the flow of energy. *Erkenntnis* 14: 219-50.

Kim, Jaegwon. 1973. Causation, nomic subsumption, and the concept of event. *Journal of Philosophy* 70: 217-36.

Lewis, David. 1999. Causation. In *Metaphysics: An Anthology*, edited by Jaegwon Kim and Ernest Sosa (Malden: Blackwell Publishing). First published in *Journal of Philosophy* 70 (1973): 556-67.

Russell, Bertrand. 1948. *Human Knowledge: Its Scope and Limits*. New York, Simon and Schuster.

Salmon, Wesley. 1998. *Causality and Explanation*. Oxford: Oxford University Press.

Endnotes

1. See Kim (1973) and Davidson (1980); Bennett (1988); and Armstrong (1997), respectively.

2. See Fair (1979); Russell (1948), Salmon (1998), and Dowe (2000); Ehring (1997), respectively.

3. Strictly speaking, this is not quite right. As I describe below, Lewis uses the term 'counterfactual' to refer to relations between the propositions that correspond to events in causal dependency relations to each other. The important point here is that Lewis's analysis of causation has putatively to do with discrete causation between events.

4. Importantly, Lewis's account is in terms of ordinary events, which are perhaps thought of as discrete entities. If the argument presented here is sound, Lewis's account need not be in terms of discrete ordinary events, and therefore nothing in his account hangs on events being taken as discrete entities. It is this failure to distinguish between discrete and continuous causation that shows Lewis's analysis is too weak.

5. Says Lewis parenthetically: "If no two events occur at exactly the same worlds–if, that is, there are no absolutely necessary connections between distinct events–we may add that this correspondence of events and propositions is one to one" (1999 [1973], 439). Notice, though, that a one-to-one correspondence relation is not required on Lewis's account.

6. Importantly, though, the converse is not the case, for reasons of transitivity: Given three events *c*, *d*, and *e*, if *d* causally depends on *c*, and *e* causally depends on *d*, it is not necessarily the case that *e* would not have occurred had *c* not occurred. Causal overdetermination cases provide apt examples: Supposed, in the actual world, Black and White both shoot Green causing Green to die, and Green's death caused the Green's obituary to be written for the following day's paper. But in a nearby possible world, Black's shot is on target while White misses; Green, however, nonetheless dies, which causes Green's obituary to be written for the following day's paper. In the actual world White's shot caused Green's obituary to be written, but as the nearby possible world shows, the writing of Green's obituary does not causally depend on White's shot. Therefore, there can be causation without causal dependence.

7. "[O]ne event is a *cause* of another iff there exists a causal chain leading from the first to the second" (Lewis 1999[1973], 440).

8. "Let *c*, *d*, *e*, ... be a finite sequence of actual particular events such that *d* depends causally on *c*, *e* on *d*, and so on throughout. This sequence is a *causal chain*" (Lewis 1999[1973], 440).

9. "Let *c* and *e* be two distinct possible particular events. Then *e depends causally* on *c* iff the family *O*(*e*), ~*O*(*e*) depends counterfactually on the family *O*(*c*), ~*O*(*c*)" (Lewis 1999[1973], 439). The statements '*O*(*e*)' and '~*O*(*e*)' read, respectively that event *e* occurs and that it is not the case that event *e* occurs. Likewise, '*O*(*c*)' and '~*O*(*c*)' mean that event *c* occurs and that it's not the case the event *c* occurs, respectively. Since *O*(*e*) is true in all and only those worlds in which *E* is

true, $O(e)$ is synonymous with E; $\sim O(e)$ is true in all only those worlds in which $\sim E$ is true and, thus, $\sim O(e)$ is synonymous with $\sim E$. For similar reasons, $O(c)$ is synonymous with C, and $\sim O(c)$ with $\sim C$. Therefore, by substitution, it can be said that e depends causally on c iff the family E, $\sim E$ depends counterfactually on the family C, $\sim C$.

10. An interesting point: One might offer the following counterexample to the counterfactual analysis of causation based on changing the middle operator: $t_1 \leq t_a \leq t_b \leq t_2$, that is, the middle operator is not '<', but '≤'. For, consider that on (CC), A at t_1 causes B at t_2 iff *(a)* A at t_1 and B at t_2 both occur, and had A at t_1 not occurred, then B at t_2 does not occur; but if $t_1 = t_a = t_b = t_2$ such that A at $t_1 = B$ at t_2, then if A occurs, then both A and B would occur, and had A not occurred, then B would not have occurred. Therefore, on (CC), every event identical with itself causes itself. I only mention this possible objection to (CC) here, without further elaboration.

11. This is the standard conception of a causal process as described in Russell (1948) and (Dowe 2000).