Asymmetry and Overdetermination in Swain's Counterfactual Theory of Causation

CHRISTOPHER HITCHCOCK
University of Pittsburgh

Hume's second definition of causation described effects as being counterfactually dependent upon their causes: one 'object' caused another "where, if the first object had not been, the second never had existed."[1] This definition lay dormant for more than two centuries before it was revived and given its best known formulation by David Lewis.[2] Several years ago, however, an alternative analysis of causation using counterfactuals was suggested by Marshall Swain.[3] I wish to examine Swain's account critically as a potential alternative to Lewis's.

Swain's analysis begins with some important definitions:[4]

(D1)   The proposition that if A were true, then C would be true is true at a world w iff:
either   (1) there are no possible A-worlds
or   (2) some A-world where C holds is closer to w than is any A-world where C does not hold.

(D3)   Where c and e are occurrent events, e depends counterfactually on c iff: if c had not occurred, then e would not have occurred.

(D4)   c and e are distinct events iff:
(1) c is not identical to e; and
(2) neither event is a constituent of the other.[5]

(D5)   Where c and e are occurrent events, e depends causally on c iff:
(1) c and e are distinct events; and
(2) if c had not occurred, then e would not have occurred.

---

[1] David Hume, *An Enquiry Concerning Human Understanding*, Section VII.
[2] David Lewis, "Causation," *Journal of Philosophy* 70 (1973): 556-67.
[3] Marshall Swain, "A Counterfactual Analysis of Event Causation," *Philosophical Studies* 34 (1978): 1-19. Swain's recent position is essentially no different from that presented in the original article.
[4] Ibid., pp. 3-5.
[5] Swain gives a more detailed account of event distinction in "Causation and Distinct Events," in Peter van Inwagen, ed., *Time and Cause*, (Dordrecht: Reidel, 1980), 155-169.

(D6)   Where $c, d_1,....., d_n$ e is a sequence of occurrent events (but not
         necessarily a temporal sequence, and where c and e may be the
         only members), this sequence is a *causal chain* iff:
         (1) d_1 depends causally on c; and
         (2) d_2 depends causally on d_1; and...
         (n) d_n depends causally on d_{n-1}; and
         (n+1) e depends causally on d_n.

Swain suggests the following as a first approximation to an analysis of
causation:

(D7)   Where c and e are specific events that occurred, c is a cause of e iff:
         there is a causal chain of occurrent events from c to e.

He argues, however, that (D7) is inadequate for two reasons. The first is
that it does not guarantee the asymmetry of causation. If c is a cause of e,
then e can not be a cause of c. (D7) seems to violate this principle. If e had
not occurred, then *some* cause of e must also have failed to occur. If c is the
cause which would fail to occur, there will be a chain of causal dependence
from e to c; so according to (D7), c caused e and e caused c. Lewis argues that
the relation of counterfactual dependence *is* asymmetric, and that causal
asymmetry follows as a corollary. According to Lewis, the closest possible
world in which e does not occur will be one in which all of the causes of e
occur anyway. In his paper 'Counterfactual Dependence and Time's
Arrow,'[6] Lewis offers criteria for judging, of two worlds w_1 and w_2, which is
closer to the actual world w. His criteria, however, do not guarantee that
w_1 will be closer to w than is w_2 if w_2 violates the laws of w and w_1 does
not. Thus, the closet possible world in which a particular event e does not
occur may (and typically will) be one in which the laws of nature of the
actual world are violated. Swain rejects this view of counterfactuals; he
argues that the candidates for being the possible world closest to w will be
ones in which the natural laws of w are not violated.[7] As a result, Swain's
relation of counterfactual dependence will not, in general, be asymmetric.
Swain tries to capture the asymmetry of causation by imposing two
additional conditions on definition (D7):[8]

(D7')   Where c and e are specific events that occurred, c is causally prior
         to e iff:
         (i)   there is a causal chain of occurrent events from c to e;

---

6 Lewis, "Counterfactual Dependence and Time"s Arrow," *Noûs* 13 (1979):
455-476.
7 See Swain, "A Counterfactual Analysis," pp. 7-10.
8 Swain, "A Counterfactual Analysis," p. 11.

(ii)   where w_1 is a world in which c occurs and e does not occur, and w is the actual world, w_1 would only have to have been different from w in the following respect: some event a (other than c) which occurs in w and upon which e depends causally in w fails to occur in w_1;

(iii)   where w_2 is a world in which e occurs and c does not occur, and w is the actual world, w_2 would have to be different from w in at least the following respects: (1) some event f (other than e) which occurs in w and upon which c depends causally in w fails to occur in w_2; and (2) some event g occurs in w_2 such that e is not causally dependent upon g in w but e is causally dependent upon g in w_2.[9]

Informally, c can be a cause of e, but not vice versa, if there is a world in which c occurred, but e didn't, which is closer to the actual world than is any in which e occurred and c didn't, where the distance between two worlds is measured by the number of events that occur in one world but not both.[10] This is intended to rule out the possibility of c and e being treated as causes of one another.

The second problem with (D7), according to Swain, is that it will fail in cases of causal overdetermination: cases where two (or more) events, each one sufficient by itself to cause e, both occur, each having an equal claim to being a cause of e. It is not true of any overdetermining cause, that had *it* not occurred, the effect would not have occurred. An example would be a man's being killed by being simultaneously shot through the heart by two bullets: had either bullet not been shot, he still would have died. Causal overdetermination must not be confused with causal *preemption*. In cases of causal preemption, two (or more) events occur, each sufficient *by itself* to cause e, but where the occurrence of one prevents the others from actually having a hand in bringing about the effect. (D7) is equipped to treat cases of causal preemption,[11] but in cases of overdetermination, it runs into trouble. Swain's strategy for tackling the problem of

---

[9] A printing error in the original article caused the last clause to misread: "some event g occurs in w_2 such that e is not causally dependent upon f in w but e is causally dependent upon f in w_2."

[10] In this informal paraphrase, I ignore difficulties in interpreting the modal expressions "would have to be different" and "would only have to have been different". Since definition (D7") already involves quantification over possible worlds, it is to be hoped that these modal expressions can be eliminated in favor of quantified expressions. Swain's intentions are not entirely clear on this point, however.

[11] See, e.g. Lewis, "Causation," p. 191; Swain, "A Counterfactual Analysis," pp. 13-4.

overdetermination rests on the following observation. Suppose two events, c and d, causally overdetermine the event e. In the actual world, neither c nor d is causally prior to e. If d had *not* occurred, however, c would have been causally prior to e by virtue of a mechanism or process which *actually* occurred. This is in contrast with the case where c is *preempted* by d: in this case, c, in d's absence, would have been causally prior to e by virtue of a mechanism which did *not actually* occur. Swain's account of overdetermination transforms this observation into an alternative set of conditions sufficient for two events to be related as cause and effect. Adding his account of overdetermination to his account of asymmetry yields Swain's full account of causation:[12]

(D7")   Where c and e are specific events that occurred, c was a cause of e iff:

Either   (A) (i), (ii), and (iii), as above in (D7'); Or

         (B) Some set of events $D=\{d_1, d_2, ..., d_n\}$ occurred (possibly having only one member) such that

               (a) If c had not occurred, and if any member d_i of D had occurred, but no other members of D had occurred, and if e had occurred anyway, then there would have been a causal chain from d_i to e consisting wholly of occurrent events, and d_i would have been causally prior to e; and

               (b) If no member of D had occurred, and if c and e had occurred anyway, then there would have been a causal chain from c to e consisting wholly of occurrent events, and c would have been causally prior to e.

Unfortunately, the seeming advantages of Swain's revisions to (D7) vanish under careful examination. I will present several examples where Swain's analysis fails to correctly diagnose the causal situation.[13] Some will exploit technical details in Swain's account. It will be argued, however, that the difficulties raised by these examples can not be circumvented by severing the relevant details from the rest of the account. Other examples will point to more general difficulties with Swain's analysis.

---

[12]Swain, "A Counterfactual Analysis," p. 16.

[13] It deserves to be mentioned that Martin Bunzl ("Causal Preemption and Counterfactuals," *Philosophical Studies* 37 (1980): 115-124) and Wayne Davis ("Swain's Counterfactual Analysis of Causation," *Philosophical Studies* 38 (1980): 169-176) have suggested counterexamples to Swain's account. I believe these counterexamples may be handled by straightforward and natural revisions, but I will not discuss this issue here.

The first telling problem is that according to Swain's analysis, there are no causes; at any rate, the existence of causes implies a restriction on possible worlds which seems unwarranted. As a preliminary, note that while condition (D7") is disjunctive, disjunct (B) is parasitic upon the condition described in disjunct (A). The clause "c would have been causally prior to e" in (D7")(B)(b) entails that there is some possible world w (not necessarily the actual world) in which c and e satisfy the relation of causal priority defined in (D7') and repeated in (D7")(A). It follows that if c does not stand in the relation of causal priority to e in *any* possible world, then both disjuncts of (D7") will fail.

Now suppose that c and e are any two occurrent events-we may suppose, to make the counterexample more vivid, that c actually causes e. If c is a cause of e, according to Swain's analysis, there must be some possible world in which c is causally prior to e. That is, there must be some world, w, with respect to which c and e satisfy condition (D7'). Clause (iii) of (D7') asserts: "where w_2 is a world in which e occurs, and c does not occur,...w_2 would have to be different from w in at least the following [respect]:...some event g occurs in w_2 such that e is not causally dependent upon g in w but e is causally dependent upon g in w_2." It is easy, however, to construct a world in which e occurs and c doesn't, which does not have such an additional causal dependence. Let w_2 be the world which differs from w in these respects: 1) c fails to occur in w; and 2) e is brought about by two *overdetermining* causes (say, d and f), neither one of which occurs in w. (See diagram 1.)

Diagram 1
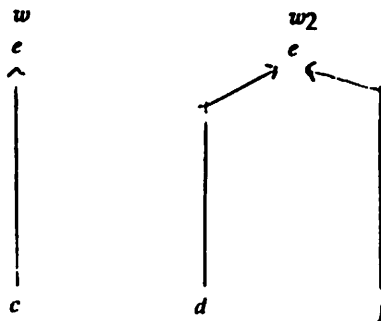
Since the events d and f *overdetermine* e, e is *causally dependent* on neither of them, so condition (D7')(iii) fails. (Recall that causal dependence, as defined in (D5), requires counterfactual dependence, which is absent in cases of causal overdetermination.) There is a similar counterexample if e is brought about by a pre-empting cause in w_2.

No argument has been offered for the conclusion that w_2 is a possible world, but it is hard to see what might disqualify it. Swain allows that an

effect may be brought about by different causes in different possible worlds, and he admits that an effect may be brought about by overdetermining causes. w_2 simply combines these two possibilities.

One possible solution to this problem immediately suggests itself: allow for the possibility of disjunctive events. In the example above, *d-or-f* could be considered as a single event. If *this* event had not occurred-if neither *d* nor *f* had occurred-in w_2, then *e* would not have occurred. With this ontological addition the counterexample vanishes. This would patch up the problem, but it would raise a new problem: Swain would owe an account of disjunctive events. Simply admitting disjunctive events without some modification of the theory would lead to undesirable results. For example, suppose my striking the cue ball on a particular occasion caused the sinking of the eight ball. Then Swain's current account would treat the disjunctive event of *my-striking-the-cue-ball-or-a-nuclear-holocaust* as a cause of the eight ball's going into the side pocket; this disjunctive event would also satisfy Swain's criterion (D7'). This is certainly odd: had the disjunctive event occurred by virtue of the occurrence of the second disjunct— the nuclear holocaust--the eight ball would have disintegrated rather than have gone into the side pocket. Thus, if Swain is to modify his account by admitting disjunctive events, he will have to make many further revisions. *Ad hominem,* Swain's choice to treat causal overdetermination by adding disjunct (B) to (D7"), rather than by admitting disjunctive events, shows an unwillingness to make these revisions.

There are, of course, many more revisions to (D7') which might be attempted. The problems uncovered by the above counterexample, however, are deep enough to raise doubts about the possibility of any successful revision. It may well be that the only revision of (D7')(iii)(2) that will save the causal phenomena is the replacement of "*e* is causally dependent upon g in w_2" with "*e* is *caused* by *g* in w_2" or some extensional equivalent. It seems at least *prima facie* likely that in order for *e* to be brought about in the absence of one of its actual (non-overdetermining) causes, it would be both necessary and sufficient to have some event be a new cause of it. Of course, to make this modification would be to flirt with vicious circularity. Swain himself foreshadows this difficulty:

> ...to imagine *e* without *c*, you need to imagine *that something prevented c from occurring and something else caused e,* but to get *c* without *e* you only need to imagine that *something prevented c from causing e.* Of course, *this* way of putting the matter neatly begs the question concerning the causal facts in this case, but the difference noted provides us with the key to capturing asymmetry.[14]

---

[14] Swain, "A Counterfactual Analysis," p. 11.

Our counterexample involving an overdetermined event in one possible world satisfies Swain's informal, circular criterion for *c* to be a cause of *e*, but not his formal, non-circular one. This suggests that the circularity is not so easily avoided.

I wish to discuss one final flaw which Swain's account of causal asymmetry shares with any counterfactual analysis of causation which disqualifies worlds which violate the actual laws of nature when evaluating counterfactuals. Such accounts rob counterfactual theories of their ability to treat cases of causal preemption. Suppose, for instance, that *c* caused *e* via *d*, preempting the causal chain from *f* to *e* via *g*. (See diagram 2.)

Diagram 2
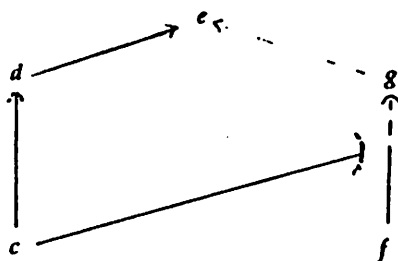
If *c* had not occurred, *f* would have caused *g*, which would have caused *e*. As it happened, however, *c* prevented *g* from occurring, thus breaking the link from *f* to *e*. The success of Swain's account depends on there being a chain of counterfactual dependence from *c* to *e*. It is not true that if *c* had not occurred, then *e* would not have occurred; it is alleged, however, that if *c* had not occurred, then *d* would not have occurred, and that had *d* not occurred, *e* would not have occurred (the causal chain from *f* to *e* having already been preempted due to *c*'s preventing *g*). The success of the latter counterfactual, however, demands that the counterfactual 'If *d* had not occurred, *c* would have occurred anyway' be true: a dangerous assumption if the actual laws of nature must hold in the closest possible world in which *d* fails to occur. *Some* cause of *d* must fail to occur in this closest possible world, and there is no guarantee that *c* will not be this disappearing cause.

Pushing the above problems aside, we may now focus our attention on the second disjunct of (D7''), which gives Swain's account of overdetermination. There is a serious flaw with this account: it grants preempted causes the status of genuine causes. Suppose that *d* caused *e*, but that *c* would have caused *e* had *d* not occurred. (See diagram 3.)

Diagram 3
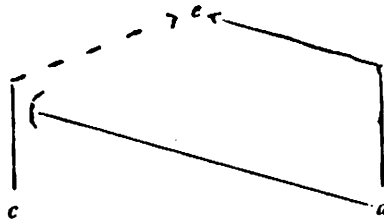
On close inspection, we see that c and e satisfy disjunct (B) of (D7'):

Consider the set $D=\{d\}$.

(a) If c had not occurred, and if d (the only member of D) had occurred, and if e had occurred anyway, then there would have been a causal chain from d to e consisting wholly of occurrent events, and d would have been causally prior to e. This follows directly from the hypothesis the d did, in fact, cause e.

(b) If no member of D had occurred, and if c and e had occurred anyway, then (i) there would have been a causal chain from c to e consisting wholly of occurrent events, and (ii) c would have been causally prior to e. Clause (ii) is satisfied by hypothesis: if d had not occurred, c would have caused e.

Swain evidently wants to prevent the analysis from going through by saying that (b)(i) is not satisfied: since c was, in actuality, preempted, there is no causal chain from c to e *consisting wholly of occurrent events*. Here Swain would be simply mistaken. There is such a causal chain, namely the sequence c, e: in the absence of d, e is causally dependent upon c, and both c and e are occurrent events. The fact that in the absence of d, the events leading from c to e did not all occur in the actual world does not affect the status of the causal chain c, e. Clearly, Swain's notion of a causal chain consisting wholly of occurrent events is a much weaker one than it first appeared: it does not come close to the intuitive notion of an occurrent causal mechanism or process which motivated Swain's account of causal overdetermination. While Swain's strategy is intuitively comprehensible, it is hard to see how it can be explicated in Swain's vocabulary of causal chains.[15]

---

[15] Peter Menzies ("Probabilistic Causation and Causal Processes: A Critique of Lewis," *Philosophy of Science* 56 (1989): 642-663) offers an objection to Lewis's account which is analogous to that raised in this paragraph.

I share Swain's discomfort with Lewis's counterfactual theory of causation. Swain's goal, a counterfactual theory of causation in which the actual laws of nature are held sacred when evaluating the possible world closest to the actual one, is an admirable one. The failure of his account to provide a sound alternative to Lewis's, however, suggests that Swain's admirable goal is also an impossible one.[16]

References

Bunzl, Martin (1980). "Causal Preemption and Counterfactuals." *Philosophical Studies* 37: 115-124.

Davis, Wayne (1980). "Swain's Counterfactual Analysis of Event Causation." *Philosophical Studies* 38: 169-76.

Hume, David (1748). *An Enquiry Concerning Human Understanding*.

Lewis, David (1973). "Causation." *Journal of Philosophy* 70' 556-67.

_____ (1979). "Counterfactual Dependence and Time's Arrow." *Noûs* 13: 455-476.

Menzies, Peter (1989). "Probabilistic Causation and Causal Processes: A Critique of Lewis." *Philosophy of Science* 56: 642-663.

Swain, Marshall (1978). "A Counterfactual Analysis of Event Causation." *Philosophical Studies* 34: 1-19.

_____ (1980). "Causation and Distinct Events," in Peter van Inwagen, ed., *Time and Cause* (Dordrecht: Reidel), pp. 155-169.