In Charles Fillmore, George Lakoff, & Robin Lakoff, eds., 1974,
Berkeley Studies in Syntax and Semantics, Volume I. Berkeley,
CA: Department of Linguistics & Institute of Human Learning,
University of California, I-1 - I-14.

Some problems in giving an adequate model-theoretic account of CAUSE[*]

Barbara Abbott
abbottb@msu.edu

David Lewis has recently proposed a counterfactual analysis of causation. The purpose of this paper is to bring up a class of cases to which the ordinary notion of causation does not seem to apply, but which would nonetheless be subsumed under it on his account, and to explore some of the difficulties in ruling out such cases using the framework in which the analysis is presented. Although I'm focusing on Lewis's treatment here I should mention that David Dowty presented independently a similar analysis in his dissertation to which the same objections, if they are valid, would apply. I would also like to thank both of these people for talking with me about these questions.

Lewis's account of causation relies on an analysis of counterfactuals which runs as follows: taking as given a weak ordering of possible worlds as nearer to the actual world depending on their higher degree of "comparative overall similarity" to it, and in which the actual world is closer to itself than any other, A □-> C ("if A were true then C would be true") is true iff either (1) there are no possible A worlds (in which case the counterfactual is vacuously true) or (2) some A-world where C holds is closer to the actual world than any A-world where C does not hold. In other words, A □-> C is nonvacuously true just in case "it takes less of a departure from actuality" to make A true along with C, than it does to make A true without C (Lewis 1973: 559f).

If c and e are actual events, and C and E the corresponding propositions that they occur, then e <u>depends causally</u> on c iff the counterfactuals

(1) a. C □-> E "If c were to occur, then e would occur"
b. ~C □-> ~E "If c hadn't occurred, then e wouldn't have occurred"

hold. To quote Lewis, "as we say it, whether e occurs or not depends on whether c occurs or not" (563). Causation itself is defined among actual events which are linked by a chain of actual events, each one causally dependent on the last. Disregarding cases in which there are more than two links in the chain, we can see that c is a cause of e whenever the corresponding counterfactual ~C □-> ~E holds.[1] To take an example, it

[*] This paper was presented at the 1974 summer LSA meeting, held in conjunction with the LSA Summer Institute at UMass, Amherst. The current version is the same as the one published in BS[3] except for a few changes in spelling and punctuation.

[1] There are 2 major differences between Dowty's and Lewis's analyses. First, Dowty's CAUSE is Lewis's causal dependence among actual events – that is it holds when and <u>only</u> when ~C □-> ~E is true (cf. Dowty 1972: 125). Lewis extends causation to a transitive relation by letting it hold whenever two events are linked by a causal chain. The difference becomes important in cases of "preemption". These are cases in which two events, say c_1 and c_2 would each by themselves have been sufficient to cause e, but where c_1

would be true to say that my alarm clock's going off was a cause of my waking up this morning if the corresponding counterfactual *If my alarm clock hadn't gone off, I wouldn't have woken up* is true. And that in turn is true if it takes less of a departure to have the alarm not go off and me keep on sleeping, than it does to have the alarm not go off and me wake up anyway.

I should perhaps mention here that Lewis states that his intention is "to capture a broad and nondiscriminatory concept of causation", and that in giving his account he is not concerned with the principles by which we may talk about causes as distinct from causal factors or conditions but is rather concerned with "the prior question of what it is to be one of the causes (nonselectively speaking)" (558f). But a case could be made that, if we reduce the notion of causation to essentially something like necessary conditionship, something which constitutes a part of our ordinary notion of a cause is left out. Recall the pair of conditions required for the relation of causal dependence between events: C $\Box\!\!\rightarrow$ E and ~C $\Box\!\!\rightarrow$ ~E. Of these the first states a relation of local sufficiency, and the second one of local necessity.[2] I think I would want to argue that the first relationship is at least as much a part of natural language causation as the second, but it is this requirement which disappears in the relation of causation as defined here. By virtue of the fact that it is automatically satisfied by any pair of occurring events, it carries no weight in the final characterization of what it is to be a cause.

The primary difficulty, it seems to me, in applying this account to the semantics of ordinary language verbs like *cause* and *bring about* comes from exactly this fact. It fails to distinguish cases in which we would say that one event caused another, from those when it would be true to say only that the first event allowed the other to occur. For instance, in the previous example my going to sleep the night before would ordinarily be thought of as something which made it possible for me to wake up. Intuitively, the difference between the relation of my going to sleep, and that of my alarm clock's ringing, to my waking up is that the second, and not the first, is regarded as locally sufficient for that event. Some further examples of sentences which are certainly deviant in this respect, and which we might even say were not true under ordinary circumstances, are:

(2) a. My lighting John's cigarette caused him to smoke it.
b. Mary's getting married caused her to become a widow.
c. The creation of Mt. Everest caused Hilary to climb it.
d. My being born caused me to come to Amherst.
e. The theft of the jewels caused the police to discover it.

---

nullifies the effect of c_2, and would thus be said to have caused e rather than c_2. In this case we do not have ~C_1 $\Box\!\!\rightarrow$ ~E (since in the absence of c_1, c_2 would have caused e); but Lewis predicts that there will be an intermediate event d, such that the pair ~C_1 $\Box\!\!\rightarrow$ ~D, and ~D $\Box\!\!\rightarrow$ ~E hold, and thus causation between c_1 and e. (Cf. Lewis 1973: 563, 567.)

The other difference is that Dowty's CAUSE-al relation is not constrained to hold only among events, while Lewis's is (although he does not reject the position that a relation of causation can hold between things other than events; cf. 558). This is an important point to bear in mind, since any application of these accounts to natural language causing (which is in fact the substance of Dowty's proposal) will have to allow things other than events as both causes and effects.

[2] These terms, and also "global necessity", below, were inspired by McCawley's felicitous dubbing of the $\Box\!\!\rightarrow$ relation as "local entailment" (McCawley 1974).

For most of these our inclination is to change the verb *caused* to something like *allowed* or *made it possible*.

Something should be said to forestall a possible objection. It may be noticed that in all of these examples the first event referred to is what might be called a globally necessary condition for the occurrence of the second event. In other words an event of the first type would be necessary for an event of the second type under any circumstances. It might then be claimed that the reason for the oddity of these sentences is not anything to do with the verb used but is rather a function of the fact that in each case the first event is entirely predictable given the occurrence of the second and that the sentences are peculiar for purely conversational reasons. This objection may be countered on two fronts. First, although the existence of something blocking light seems globally necessary for the existence of a shadow, a sentence like

(3) The shadow on the wall was caused by the presence of something blocking the light.

is not at all anomalous. This is so, I claim, because the presence of something blocking the light is also sufficient to cast a shadow. On the other hand one can find examples of events only locally necessary for the occurrence of others, and thus not totally predictable given the occurrence of the second, and yet which show the same sort of anomaly which we find in the first set of examples. For instance, in the situation in which I have lost my car keys, it might be necessary for me to hotwire my car before driving into town. Hotwiring my car is here a necessary yet not predictable precursor to my driving it, but

(4) My hotwiring my car caused me to drive into town.

is odd in the same way that the sentences of (2) are odd, because, again, hotwiring my car is not ordinarily sufficient to bring about my driving into town.

In support of my claim, notice that examples like those first given may be made more suitable by devising contexts in which the first event becomes locally sufficient for the occurrence of the second. For instance, should it be the case that someone is holding John down and pumping his chest in and out and someone else is holding a cigarette firmly between his lips while keeping his nostrils shut, and then I light the cigarette, it would perhaps be true to say that my lighting the cigarette brought about John's smoking it. Similarly if John was a person who was on the verge of death, we might say that Mary made herself a widow by marrying John. On the other hand, Hilary's declaration that he climbed Mt. Everest because it was there is so effective because it implies that the mere fact of the mountain's existence was sufficient for him to attempt the climb.

In looking for a means to rule out as causes events which are not locally sufficient for the caused event, the logical first choice might be to add to the requirements for causation the condition in (5).

(5) $\sim E \ \Box\!\!\rightarrow \sim C$

This would seem to have the desired effect with respect to the anomalous sentences given above since in each case it would seem to be less of a departure from the world in which these events occur to get rid of the effect in question by some other means than by eliminating the event in subject position.

The addition of this condition solves another problem which arises with the original analysis. (The problem itself I believe originates in Hart & Honoré's *Causation in the Law* (pp. 35f). It was raised in this context by Jim McCawley in the paper cited in footnote 2.) We might want to say that

(6) The roses died because the gardener didn't water them.

is true, and that

(7) The roses died because Henry Kissinger didn't water them.

is not. And yet at the same time it seems to be true that if either of these people had watered the roses they would have lived. On the other hand a world in which the roses live and the gardener has watered them is likely to be closer to ours than any in which they live and the gardener hasn't watered them., while the same would not hold true for Henry Kissinger.

At this point we should perhaps consider the possibility of making do with (5) alone.[3] A generous translation of (5) might be

(8) If e hadn't occurred, that would have meant that c hadn't occurred.

or, an even more sympathetic interpretation suggested by Lewis (personal communication)

(9) If e hadn't occurred, that would have been because c hadn't occurred.

But logic is notoriously ungenerous, and what (5) in fact says is

(10) If e had not occurred, then c would not have occurred.

Exactly the same as (1b) with c and e reversed. What this means is that any pair of propositions satisfying (1b) will also satisfy (5), except that we would now be calling the cause the effect and the effect the cause. Thus in our first example, my waking up would be accounted a cause of my alarm clock's going off (as well as of my having gone to sleep the night before) and this is surely an intolerable result.

Unless some mechanism can be found for preventing these consequences, we can for the moment consider (5) viable only in conjunction with (1b), but while the addition of (5) appears to solve the gardener problem, it has serious side effects which must be taken into account. The first of these (one which was pointed out by Lewis (566)) is that cause and effect become thereby indistinguishable. Furthermore, Lewis remarks, any attempt to distinguish them through the use of temporal priority is useless against another

[3] This was in fact suggested by McCawley in "Remarks on what can cause what".

problem which runs as follows. Suppose event c causes both e and f, and that e precedes f. With the conditions as stated there will be no way to tell that it was c, and not e, that caused f. A nice example of Lewis's which illustrates this case is one in which c is the falling of a hatchet, e the shadow of the hatchet, and f the decapitation of a victim.

It may be that this problem is solvable through the use of some means other than temporal priority to distinguish cause from effect. Von Wright, for instance, in *Explanation and Understanding*, argues that cause may be distinguished from effect by a "requirement of manipulability" – "p is a cause relative to q, and q an effect relative to p, if and only if by doing p we could bring about q or by suppressing p we could remove q or prevent it from happening" (70). On the other hand Reichenbach, in *The Rise of Scientific Philosophy*, suggests that cause and effect are naturally distinguished in a number of ways – that they are generally irreversible processes, for instance; ones which proceed from order to disorder, such as the mixing of coffee and cream (146ff).

However, even if one of these suggestions, or some other means, could be put to work to separate a symmetrically conditioned cause and effect, condition (5) presents additional difficulties of its own. The first of these problems is that the condition as stated would seem to single out at most only one cause per effect; but what is worse, it makes its selection on what appear to be irrelevant grounds. To see how this works, take the example of a car accident, caused, we might say, in part by an ice storm a few hours before, and in part by the running out into the road of a dog at an inopportune moment. Without either of these two events, we can safely assume, the accident would not have happened, so they both meet condition (1b). When we come to (5) however, what we are looking for is an event which meets the condition that it not have occurred in all the closest alternative worlds in which the accident itself did not occur. Now among these non-accident worlds, one in which the dog failed to run out into the road at just the time he did (but in which the ice storm still occurred) would seem to be closer on any intuitive grounds of comparative overall similarity to the original world than any in which the ice storm did not occur, or occurred at a different time. Hence the ice storm, by virtue of its having made a bigger splash so to speak in the actual world, gets ruled out as a cause of the accident.

In fact condition (5) would seem to suffer a pathological breakdown in a situation in which, among the contestants for non-existence in the closest worlds in which the effect did not occur, no one of them was such that its removal from the original world would make significantly less of a difference to that world than the removal of any of the others. For in that case no event would meet the condition that it not have occurred in all the closest alternatives in which the effect did not occur.[4]

[4] It might be thought that this problem could be gotten around by allowing here the conjunction of two such events to be called a cause. In a sense this is what we want to say in such a situation – that a and b jointly caused c, though neither would have been sufficient on its own. This way out is not without attendant problems. First we would be clearly deviating from Lewis's analysis since the conjunction of two events is not in general itself an event. But more importantly, the question would then arise as to how to exclude the conjunction of a real cause and any other irrelevant event (which meets one requirement, noted below) as a cause of an effect. To see that this problem arises, note that if ~E $\Box\!\!\rightarrow$ ~C holds, then ~E $\Box\!\!\rightarrow$ ~(C & D) will hold for any arbitrary event d. Furthermore if ~C $\Box\!\!\rightarrow$ ~E holds, then ~(C & D) $\Box\!\!\rightarrow$ ~E will hold for any arbitrary d as long as its removal from the actual world makes more of a difference than the removal of c.

Furthermore, if we continue to restrict causation to a relation between events, condition (5) would refuse to grant us any cause at all in another sort of circumstance, and for similar reasons. That is, no event could be called the cause of another if there is some third, non-occurring event whose occurrence would have prevented the effect in question while making less of a difference to the original world than the non-occurrence of the would-be cause. For example; we might want to say that Watergate was the cause of Nixon's downfall. In order to meet condition (5) it would have to be the case that in the closest worlds in which Nixon's reputation was untarnished, Watergate (meaning the infiltration of the Democratic National Headquarters) did not occur. Yet we might also want to say that had the night watchman been called outside just before he would have seen the men or the bug or whatever it was he saw, the whole thing would never have been made public and Nixon would have been safe. Yet if we accept condition (5) we can't have both of these things being true. This is because among those worlds in which Nixon stands untouched, the one in which the night watchman was called outside and failed to see the men would surely take less of a departure from our world than one in which the goings on at Watergate never occurred in the first place.

These problems are related, I believe, to the inadequacy of ~E $\Box\!\!\rightarrow$ ~C as a substitute for C $\Box\!\!\rightarrow$ E, which is reflected in the difficulty of finding a proper translation for the former conditional which says what we want it to mean – that c was a locally sufficient condition for e. Both (11a) and (11b)

(11) a. If my alarm clock hadn't gone off, I wouldn't have woken up.
b. If my alarm clock had gone off, I would have woken up.

are perfectly natural conditional statements and engender clear intuitions as to their truth or falsity. But

(12) If I hadn't woken up, my alarm clock wouldn't have gone off.

by virtue of its unnaturalness as a conditional, affords us no such clear intuitions. In a sense, use of ~E $\Box\!\!\rightarrow$ ~C as a condition for causation seems a perversion of the original counterfactual analysis.

There is a quotation from Hume (from *An Enquiry concerning Human Understanding*, Section VII), cited by Lewis at the start of his paper, which runs:

> We may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second had never existed*. (Lewis 1973: 556; italics in original.)

Of the two clauses, the first makes a cause a sufficient condition, the second a necessary one. As Hart & Honoré point out, the two definitions fall together if we make the simplifying but contrary to fact assumption that any event has one and only one cause (122). But, as they also note, it was one of Mill's insights that the first object in Hume's definition is in fact a complex one, comprised of numerous events and conditions which are only together sufficient to produce the second, and it was Mill's position that "we

have philosophically speaking no right to give the name of cause to one of them exclusively of the others" (Mill *A System of Logic*: Book III, Ch. V, § 3; quoted in Hart & Honoré 1959: 16).

As far as natural language goes we seem to fall somewhere between the two positions. When we say X caused Y we seem to imply that X by itself was sufficient for the job. But on the other hand we also say things like *Y was caused largely by X*, or *Y was in part due to X*, perhaps not so much because we are at these times adhering to Mill's doctrine, but because we are unable in those circumstances to find any one event or condition which appears to us to be the sole sufficient one. But in trying to characterize formally this ordinary notion we seem forced to adopt one of two extreme positions. Either we call any necessary condition for an event the cause of it, in which case we wind up calling, for instance, the event of our being born a cause of everything we do; or we add something like a condition of local sufficiency, in which case we not only have the problem of formally distinguishing cause from effect, but also we run into the difficulties outlined above.

References

Dowty, David. 1972. *Studies in the Logic of Verb Aspect and Time Reference in English*. Austin, TX: University of Texas dissertation, Ch. 5.

Hart, H.L.A. & A. M. Honoré. 1959. *Causation in the Law*. Oxford: Clarendon Press.

Hume, David. 1777. *An Enquiry concerning Human Understanding*. London.

Lewis, David. 1973. Causation. *Journal of Philosophy* 70:17, 556-567.

McCawley, James D. 1974. Remarks on what can cause what. Paper read at the USC Causative Festival, 5/2/74.

Mill, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive, being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*. London: John W. Parker.

Reichenbach, Hans. 1951. *The Rise of Scientific Philosophy*. Berkeley, CA: University of California Press. Chs. 9-10.

Wright, G.H. von. 1971. *Explanation and Understanding*. Ithaca, NY: Cornell University Press.