6*

Independence and Counterfactual Dependence

After briefly discussing two other counterfactual theories of causation, this chapter formulates the account of similarity among possible worlds employed in chapter 6 and proves the claims it makes.

6.1* Mackie's Counterfactual Theory

John Mackie argued,

> if on a particular occasion *A*'s doing *X* is causally related to *B*'s doing *Y*, and if they had not been so related but things had otherwise been as far as possible as they were, *A* would still have been doing *X* but *B* would (or might) not have been doing *Y*, then *A*'s doing *X* is conditionally and causally prior to *B*'s doing *Y*. (1979, p. 24)

For example, suppose one breaks the connection between a car's engine and its wheels: the engine continues turning while the wheels stop.

Let us call the event of *A*'s doing *X* on the particular occasion "*a*" and the event of *B*'s doing *Y* on the occasion "*b*." Suppose that in the closest possible world in which some event *c* failed to occur, *a* would not be causally connected to *b*. If the causal connection between *a* and *b* depends on the existence of *c*, then *c* must be a cause of at least one of *a* or *b*. On Mackie's view *a* is causally prior to *b* if and only if, in the absence of some minimal difference *c*, *a* still occurs, but *b* does not (*B* does not do *Y*). So *c* causes *b* only. Mackie thus indirectly assumes that there are causes of *b* that are not causes of *a*, and this assumption is, of course, an immediate implication of both **CP** and agency views. One can thus capture Mackie's intentions without introducing the further complications raised by such counterfactuals. These complications are nevertheless enlightening. Mackie has pointed out a further feature of causal priority, which is neatly explained by **CP** and $\mathbf{AT}_{\mathbf{g}}$.

6.2* Swain's Theory of Causal Asymmetry

Although invoking last-minute miracles neatly permits one to deny the counterfactual dependence of causes on effects, Marshall Swain suggests that invoking miracles also leads one mistakenly to deny the counterfactual

dependence of effects on causes. A world with a "pure deletion miracle" which "deletes" a cause "but which leaves the *rest* of the actual world entirely unchanged" would apparently be extremely similar to the actual world (1978, p. 9). If such a world is more similar than one in which both the cause and its effects are missing, then effects do not depend on their causes.[1]

Swain proposes the following alternative to Lewis's theory:

D7N Where *c* and *e* are specific events that occurred, *c* is a cause of *e* iff:

1. there is a chain of occurrent events from *c* to *e*;
2. 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 cause *a* (other than *c*) which occurs in *w* and upon which *e* depends causally in *w* fails to occur in w_1;
3. 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.[2]

In Swain's account, the laws of similar possible worlds such as w_1 and w_2 are the laws of the actual world, *w*. So for events to fail to occur or for new events to occur, the past might have to be drastically different. Swain accepts Lewis's notion of causal dependence: *e* causally depends on *c* if and only if *c* and *e* are distinct and if *c* had not occurred, then *e* would not have occurred. But he denies that causal dependence is generally asymmetric. If, as Swain holds, the laws of the most similar possible worlds are the laws of the actual world, then causes will often causally depend on their effects. Thus on Swain's account, the asymmetry of causation does not derive from an asymmetry of causal dependence. It rests instead on the fact that worlds in which *c* occurs without *e* require only that some other cause or causal condition of *e* not occur, while to have *e* without *c*, it must be the case both that some cause *f* of *c* fails to occur and that some other event *g* occurs, which causes *e* in the absence of *c*.

Swain's account supposes that whenever *c* causes *e* there is some distinct cause *a* of *e* that is not itself causally dependent on *c*. Otherwise there would be no easy way to have *c* without *e*. Similarly when Swain discusses effects of a common cause, he requires that each have its own distinct and independent causes (1978, p. 12). So Swain's account presupposes something like the independence condition **I**, and unless Swain would take issue

[1] This criticism of Lewis preceded the publication of Lewis's "Counterfactual Dependence and Time's Arrow" (1979) and seems to be answered by Lewis's discussion there of the asymmetry of miracles and overdetermination. For other criticism of Swain's view see Davis (1980).

[2] 1978, p. 11. Swain's footnote (after "only") is omitted. The last two occurrences of "*g*" in the quotation are misprinted as "*f*" in the printed text.

with the connection principle or transitivity, Swain is thus committed to the independence theory, **CP**. Once committed to **CP**, there is little reason to pursue a counterfactual theory as an *alternative*.

Swain's revision of Lewis's theory is problematic. There is no reason why the causal structure might not be such that the absence of *f* required in the second clause of **D7N** in order for *c* to be absent might not itself be sufficient to bring about *e*.[3] More generally, one might state the following worry: **D7N** replaces Lewis's claims that non-*e* possible worlds with *c* are more similar to the actual world than non-*e* possible worlds without *c*. **D7N** compares numbers of changes in the immediate vicinity of *c* and *e* rather than overall similarity. Second, rather than comparing non-*e* possible worlds, **D7N** compares *e*-and-non-*c* possible worlds to *c*-and-non-*e* possible worlds. These two changes raise two questions: What does the number of different occurrences or nonoccurrences in the immediate vicinity of *c* and *e* have to do with causal asymmetry? Second, what is the relevance of a comparison between *c*-and-non-*e* possible worlds and *e*-and-non-*c* possible worlds? Until these two question are answered, the account seems arbitrary.

6.3* An Alternative Account of Similarity Among Possible Worlds

The argument in chapter 6 relies on the following general principle concerning the similarity among possible worlds:

SIM (*Similarity among possible worlds*)

1. *Worlds with miracles are not the most similar*. For any event *b* there are non-*b* possible worlds without at least one of *b*'s causes that are at least as close to the actual world as are any non-*b* possible worlds in which all of *b*'s causes occur.

2. *It doesn't matter which cause is responsible*. For any event *b*, if *a* and *c* are any two causes of *b* that are causally and counterfactually independent of one another, there will be non-*b* possible worlds in which *a* does not occur and *c* does occur that are just as close to the actual world as are any non-*b* possible worlds with *a* and without *c*, and there will be non-*b* possible worlds without *a* and with *c* that are just as close to the actual world as are any non-*b* possible worlds without both *a* and *c*.[4]

3. *The fewer the irrelevant differences in events, the more similar the world*. For

[3] Such a case would involve a multiple connection between *f* and *c* – one connection via *c* and another via some other chain of consequences of *f* not appearing. As we will see below, multiple connections create problems for Lewis's theory, too.

[4] On the grounds that the fewer the differences, the more similar the worlds, one might question whether non-*b* worlds without both *a* and *c* can be just as similar to the actual world as are non-*b* worlds without just one of these causes. Those who find this plausible can change this clause of **SIM**. The arguments in this chapter, including the proofs of the three theorems employing **SIM** as a premise, go through just the same. If one accepted the reasoning that the fewer the differences, the more similar the worlds, one would have to predict in figure 6.4 that x_5 has one of two possible values instead of any value between the two values. It is for this reason that I prefer the present formulation.

any event *b*, consider two possible events *e* and *f* that are not causally connected to *b*, where *e* occurs and *f* does not. Then there are non-*b* possible worlds with *e* that are more similar to the actual world than any non-*b* possible worlds without *e*, and there are non-*b* worlds without *f* that are more similar to the actual world than any non-*b* worlds with *f*.

4. *The fewer the irrelevant differences in laws, the more similar the world*. For any event *b* there is a non-*b* possible world in which all other laws, apart from those relating *b* to its causes or the causes of *b* to one another, are the same as the actual world that is more similar to the actual world than is any non-*b* possible world which differs with respect to some such laws.

The first clause in **SIM** denies that possible worlds with miracles between *b* and its causes are more similar to the actual world than are possible worlds where the causal relations between *b* and its causes hold and the miracle, if any, comes earlier. It does not insist that laws must be held sacrosanct. Lewis rejects (1), but his main reason seems to be that he cannot account for the asymmetry of causation unless he locates miracles as late as possible. The second clause maintains that it is equally easy to get rid of *b* by getting rid of any combination of its independent causes. This clause is neither asserted nor denied by Lewis. The third clause says that unrelated changes detract from similarity, while the fourth says that additional differences in laws detract from similarity. Lewis would endorse the last two clauses of **SIM**. **SIM** does not offer comprehensive rules for comparing the closeness of possible worlds. Only the first clause of **SIM** is incompatible with Lewis's account.

Unlike the first, third, and fourth clauses of **SIM**, the second is not plausible. Its difficulties are discussed above in connection with the example of George jumping off the Brooklyn Bridge (p. 116). The difficulties concerning similarity are just as serious for Lewis's version of the theory as they are for the revision I am exploring. I think that difficulties in specifying a plausible relation of comparative overall similarity among possible worlds give one good reason to avoid counterfactual theories altogether.

6.4* Independence and Counterfactual Dependence

To derive implications concerning counterfactuals from the independence theory of chapter 4, one needs one further condition:

CDCC (*Counterfactual dependence implies causal connection*) If *a* and *b* are distinct events and *b* counterfactually depends on *a*, then *a* and *b* are causally connected.

CDCC states the metaphysical principle that counterfactual dependence among distinct events is always causal or nomological. This principle is not obvious, but it is widely accepted, and Lewis can have no objection to it, since it is implied by his theory. Given the links between causal connection

and probabilistic dependence, **CDCC** implies that counterfactual dependencies between individual events will typically be reflected in probabilistic dependencies in the circumstances between events of the relevant kind. One can then prove:

Theorem 6.1: SIM, **CDCC**, and **I** imply that individual causes will not be counterfactually dependent on individual effects and effects of a common cause will not be counterfactually dependent on one another.

Proof: **I** says that for every cause a of every event b there will be some other cause c of b causally independent of a. Given **CDCC**, a and c will also be counterfactually independent. Clause 2 of **SIM** then implies that for any non-b world without a, there will be a non-b world without c that is at least as similar to the actual world. So individual causes will not be counterfactually dependent on their effects. **I** also implies that if b_1 and b_2 are effects of a common cause a, then b_1 has a cause a_1 that is independent of any cause of b_2, and b_2 has a cause a_2 that is independent of any cause of b_1. Given **CDCC** and **SIM** there will be non-b_1 possible worlds without a_1 that are just as close to the actual world as are non-b_1 possible worlds without a, and in those possible worlds b_2 will still occur. So b_2 does not counterfactually depend on b_1. Given **SIM** there will also be non-b_2 possible worlds without a_2 that are just as close to the actual world as are non-b_2 possible worlds without a, and in those worlds b_1 will still occur. So b_1 is not counterfactually dependent on b_2.

Theorem 6.2: CDCC, **SIM**, **I**, and **CC** imply that if b counterfactually depends on a, then a causes b.

Proof: Suppose b counterfactually depends on a. Then (by **CDCC**) a and b are causally connected. **I** and **SIM** imply that b does not cause a and that a and b are not causally connected only as effects of a common cause. By **CC** it follows that a causes b.

Artificial event fusions create difficulties for the claim that individual causes are not counterfactually dependent on their effects. Suppose that determinism is true and that $a_1, \ldots, a_n$ cause b. If b were not to occur, while $a_1, \ldots, a_{n-1}$ occurred, then if miracles between the cause and effect under consideration are not allowed, the other cause, a_n, must have failed to occur. This does not, of course, show that if b had failed to occur a_n would not have occurred. It shows instead that if a_1 through a_{n-1} occurred and b had failed to occur, then a_n would not have occurred. Let e be the event that occurs whenever a_1 through a_{n-1} and b all occur. So when a_1 through a_{n-1} occur and b does not, then e does not occur. Then it might appear that a_n is counterfactually dependent on e, and if counterfactual dependence is sufficient for causation, one will be led to the false conclusion that e causes a_n. This claim is problematic, since there are other ways that e might fail to occur, but to give a general answer to this objection requires that one say something about artificial events such as e. In chapter 13 I shall attempt to say something about when event fusions can stand in causal relations and how their causal relations depend on the relations among natural tropes. At this point, however, I am concerned only with relations among natural

events or tropes, and *e* is not a natural event.

6.5* The Asymmetry of Overdetermination

Qualms about **I** do not give one reason to prefer Lewis's version of his own theory, because **I** is implicit in the view that worlds with small miracles are closest, and because it is hard to explain the asymmetry of overdetermination, which Lewis's theory requires, if **I** is not true. Ordinary overdetermination, like that discussed in §3.5 and in §13.1, involves the existence of multiple minimal sufficient conditions. The overdetermination that Lewis is concerned with involves individual conjuncts. Let us then say that a trope *e* is *determined* by *f* if and only if *f* is sufficient in the circumstances for *e*. *f* is sufficient in the circumstances for *e* if *e* is necessary in the circumstances for *f*. Necessity in the circumstances is defined by **DC** (p. 43). *e* is overdetermined by *f* and *g* if and only if *f* and *g* are both sufficient in the circumstances for *f*.

Consider then the following strong restatement of Lewis's asymmetry of overdetermination claim:

> **AOD** (*The asymmetry of overdetermination*) If causation is deterministic, then (1) events will be determined by a great many of their (natural) effects, and (2) events will be not be determined by any of their (natural) causes.

The following theorem shows that **AOD** follows from **I** when causation is deterministic and there is no preemption or ordinary overdetermination.

Theorem 6.3: If there is no preemption or ordinary overdetermination (overdetermination by conjunctions of natural causes), then **DC** and **I** entail **AOD**.

Proof: Given **DC** and the absence of ordinary preemption or overdetermination, the conjunction of (the properties of) the cause tropes is necessary in the circumstances for the effect trope, and so each cause is determined by each of its effects. Suppose that *a* causes *b*. By **I** there will another cause *f* of *b* that is causally independent of *a*. Since *a* will not be necessary in the circumstances for *f* nor vice versa, *a* by itself – that is, without *f* – will not be sufficient in the circumstances for *b*. Effects will not be determined by any of their individual causes.

Individual causes are (given the absence of alternative sufficient conditions) necessary, but not sufficient for individual effects, so individual effects are sufficient for individual causes. Since there are cases of preemption and ordinary determination, it is not true that events are always overdetermined by their effects, but since preemption and ordinary overdetermination are rare, it is usually the case that events are overdetermined by their effects, while they are never overdetermined by their causes. **I** is only sufficient for the asymmetry of overdetermination, not necessary. Only the much weaker condition that not all of the causes of an event determine one another is

necessary. If **I** were false, the separate causes of an event would come closer and closer to overdetermining it as the lawful connections among the separate causes grew tighter. Although **I** is not necessary for the asymmetry of overdetermination, it is difficult to see why the asymmetry of overdetermination should be true if **I** were not true.

6.6* Proof of a Restricted Version of L

As argued in §6.6, causation does not imply counterfactual dependence when there are multiple connections between cause and effect. So the necessary condition that **L** states and consequently **L** itself are false. But one can prove:

Theorem 6.4: Given **CC**, **DC**, **SIM**, and no multiple connections – that is, if *a* had not occurred, no cause of *a* would have been a cause of *b* –, if *a* causes *b*, then *b* is counterfactually dependent on *a*.

Proof: Suppose that *a* causes *b* and there is no preemption or overdetermination. Then, by **DC** *a* is necessary in the circumstances for *b*. Suppose (counterfactually) that *a* does not occur, and consider the possible worlds in which *b* occurs anyway. In some worlds *b* occurs miraculously, but these will be very unlike the actual world. In others *b* occurs because of some causes. Those causes cannot be preempted actual causes, because by assumption there is no preemption. They cannot be a causes of *a*, since there are no multiple connections. So whatever causes *b* must be some new occurrence or nonoccurrence of something that does not cause *a* or *b* in the actual world, but which, as a consequence of some new law, causes *b* in this possible world. By clauses 3 and 4 of **SIM**, possible worlds with *b* will be less similar to the actual world than are some possible worlds without *b*. So *b* is counterfactually dependent on *a*.

Theorems 6.2 and 6.4 obviously imply:

Theorem 6.5: **CC**, **I**, **DC**, **SIM**, and **CDCC** entail **L** restricted to circumstances in which there are no multiple connections.

I have argued that **SIM** is more plausible than Lewis's account of similarity among possible worlds. **CDCC** is as plausible as any claim I know of relating causation to counterfactuals and unobjectionable to counterfactual theorists such as Lewis. **DC** merely restricts the argument to deterministic relations between cause and effect. So theorem 6.5 shows that the independence theory of chapter 4 explains what is true about the counterfactual theory.

Given the precise formulation of **SIM**, **L** (restricted to circumstances without multiple connections) does not directly imply or presuppose the truth of **I**. Suppose some event *e* has only one cause *c* or only two causes c_1 and c_2 that are causally connected. One can still deny the counterfactuals "if *e* had not occurred *c* would not have occurred" or "if *e* had not occurred c_1 would not have occurred" on the grounds that possible worlds in which *e*

fails to occur as the result of a miracle need be no less similar than are possible worlds in which the causes of *e* failed to occur. If, on the other hand, one maintained that worlds with the same laws relating events to their causes as the actual world are *more similar* (not merely no less similar) to the actual world than are worlds with miracles, then **L** would imply **I**. One also needs something very like **I** in order to defend Lewis's asymmetries of miracles and of overdetermination.