Chapter 4: Regularity reconsidered

§1 Overview

The critique of counterfactual analyses carried out in the last chapter pointed to a conclusion much more interesting than the claim that such analyses face apparently insuperable obstacles. It suggested, in addition, that the reason for their failure is that they can, at best, succeed only at explicating the nature of one kind of causal relation. There is another kind of causal relation—what I have termed "production"—that needs a very different sort of analysis. This chapter begins the project of providing one.

The analysis I will offer builds on two key ideas, of which for now I will provide just a rough and intuitive statement. The first idea, which I call the "tracing thesis", says that causes are connected up to their effects by way of intermediate causes, so that if we begin with some event *e*, we can find its causes by "tracing back" through less and less proximate causes. As we will see, there is more to this idea than just the claim that causation is transitive (although that is part of it). The second key idea, which I borrow from Mackie's 1967 analysis, is that what is distinctive of the causes of some event is that they collectively suffice for it, in a very specific sense that will take some work to elucidate. I call this the "sufficiency thesis".

It is the connection to Mackie that recommends the name "regularity analysis" for the kind of account I will develop. That name is also guaranteed to raise hackles. Haven't we learned long since that regularity analyses of causation are hopeless? Granted, the news hasn't reached all corners of philosophy, but among those who work on the metaphysics of causation a sentiment that Lewis gave pithy expression to 30 years ago is surely well entrenched:

> It remains to be seen whether any regularity analysis can succeed in distinguishing genuine causes from effects, epiphenomena, and preempted potential causes—and whether it can succeed without falling victim to worse problems, without piling on the epicycles, and without departing from the fundamental idea that causation is instantiation of regularities. I have no proof that regularity analyses are beyond repair, nor any space to review the repairs that have been tried. Suffice it to say that the prospects look dark. I think it is time to give up and try something else. (1973, p. 160)

I think, rather, that it is time to turn back the clock, for a regularity analysis broadly in the spirit of Mackie's own—but, as we will see, refined and updated in several important respects—can succeed admirably in overcoming the problems Lewis mentions, as well as others. Granted, it will depart significantly from the allegedly "fundamental" idea that causation is the instantiation of regularities—but then again, so did Mackie's own, and at any rate there is little interest in quibbling over the propriety of a name. Constructing this analysis will require a lot of work in shaping the sufficiency thesis, and then a bit more work in order to properly incorporate the tracing thesis. First I will sketch its broad outlines, making use of a pair of idealizations. One idealization is that the events we are concerned with are all instantaneous. The second idealization is that time passes in discrete moments, so that for each time t, there is a most recent time that precedes it. Thus for any event there is a well defined sense in which some other events are its *immediate* causes. We use the sufficiency thesis to pick out this relation of immediate causation: very roughly, the immediate causes of event *e* are those events immediately preceding it that collectively suffice for it, but that do so non-redundantly. We then appeal to the tracing thesis to define causation more generally as the ancestral of immediate causation.

Naturally, relaxing these idealizations leads to complications. I will wait until much later in this chapter to relax the first, but we must abandon the second right away, since holding onto it makes life a little bit too easy. Let me briefly indicate why. Once we recognize that there simply are no events that immediately precede *e*, we must also recognize the need for a relation of sufficiency that some of *e*'s causes can bear to it even though they only help produce *e* by way of causal intermediates. Having provided such a relation, we must prepare for the possibility that some events might suffice for *e*, even though they fail to suffice for any intermediate events that themselves suffice for *e*. We will need to figure out a principled way to deal with this sort of possibility. But more of these matters later; what is at issue will become much clearer once we have an account of sufficiency on the table.

Some last caveats before we turn to developing that account. Several important topics will be raised in this chapter—but not addressed at all adequately—and I want here to flag them. First, the analysis I will offer does not handle certain stubborn varieties of causal pre-emption. The next chapter provides a detailed discussion of these varieties, with an eye towards diagnosing what exactly makes the problem cases tick. That discussion will set the stage for chapter 6, where I will argue that the key to overcoming the problems posed by pre-emption is to recognize that the causal structure of a process is intrinsic to it. I will develop this thesis in a way that will allow for a natural extension of the regularity analysis presented in this chapter, an extension that will enable it to cleanly handle the stubborn cases of pre-emption.

Second, no provision is made in this chapter for so-called "fine-grained" causation—as, for example, when we say that some event *e* happened not because *c* happened, but rather because *c* happened in a particular way. (Billy got upset not because Suzy yelled at him, but because she yelled at him so loudly, etc.) How best to account for causal phenomena such as this is a problem I will take up in chapter 7.

Third, any account of causation that makes use of claims to the effect that such-and-such events do or do not occur—as counterfactual analyses typically do, and as my regularity analysis will—owes us an account of the content of these claims. Put another way, such an account of causation needs to be supplemented by an account of the occurrence-conditions for the events that it takes to be the fundamental causal relata. Many of the issues involved in providing such an account I will also postpone until chapter 7.

Fourth, my account of causation will build in the claim that causation is transitive, a claim that has come under heavy fire in the recent literature. Responding to these attacks, and taking up other questions involving the role of transitivity, will be the business of chapter 8.

Fifth, my account of production does not merely fail to handle, but positively rules out causation by omission; it is also far from clear how it can be extended to cover prevention, particularly if this is understood to be causation *of* omission. Intuitively, that is as it should be: whatever we mean by "produce" or "generate", it is not the sort of thing that a mere absence can

accomplish, nor the sort of thing that should result in a mere absence. But it is emphatically not that I think that there is no such thing as causation by omission or prevention. What these might amount to is a topic that chapter 9 (co-written with Sarah McGrath) will take up.

Sixth, my account makes the asymmetry of causation parasitic on the asymmetry of time, a feature many will find disappointing. And perhaps worse than disappointing: it might seem that *backwards* causation is perfectly possible, and so wrongly ruled out a priori by my account. I consider such worries overblown, and explain why in chapter 12.

Finally, I will completely ignore for the time being the question of whether there is any context sensitivity in our causal judgments, and if so how best to account for it. In fact, not only do I think that there is such sensitivity, but I think that shifts in context can sometimes make the difference between whether we count *c* a producer of *e*, or merely say that *e* depends on *c*. These and other complex issues associated with context sensitivity will be dealt with in chapter 10.

Enough bracketing. Let us now take up the task of formulating an adequate account of sufficiency that will serve the needs of our regularity analysis.

§2 The sufficiency thesis

§2.1 Mackie's analysis

As a starting point for discussion, let me sketch Mackie's analysis of causation. (But, since this is not an exercise in Mackie exegesis, the sketch will be fairly, well, sketchy.) Mackie takes a cause of an effect E to be, in the first instance, something that is an insufficient but necessary part of an unnecessary but sufficient condition for E—for short, an "INUS condition" for E. That is, the cause is not itself sufficient for the effect, but is a necessary part of a more comprehensive "condition" that *is* sufficient (though not necessary). Note that I do not use the lower case, italicized "*e*" for the effect, since Mackie does not—at least, not obviously—take causes and effects to be *events*. That observation points to the first of a number of questions that clamor for attention, most prominently these: What sort of thing is this "something"? An object? Fact? Property? Event? Mackie's talk of "conditions" does not exactly clarify matters. What does

"necessary" mean? What does "sufficient" mean? Why must a cause be insufficient for its effect, and why must the thing of which it is a part be unnecessary for that effect?

I will take up these questions shortly, but I want to make sure at the outset that we do not lose sight of the central idea, which, for all the obscurity of Mackie's presentation, is nevertheless worth pursuing. I take it to be this: The distinguishing feature of a set of causes of E is that they collectively suffice for E, but in a way that makes each constituent element essential; that is, no proper subset of these causes would suffice for E. (This claim, by the way, answers the question of what the "N" in "INUS" means.) That formulation still leaves plenty of questions unanswered—as of course it should, given that it is only intended to present the guiding idea.

I will now try to develop this idea in some detail, by cleaning up Mackie's own analysis. This clean-up will proceed in several steps.

§2.2 "INUS" vs. "NS"

Mackie's talk of "insufficient" and "unnecessary" is, well, unnecessary.[1] He has in mind, I think, a typical case of causation, where a cause does not suffice all by itself for its effect (hence the "I" in "INUS"), and where a set of causes is not necessary for the effect, in the sense that *other* causes could (consistently with the laws, say) have brought that effect about (hence the "U"). Mackie of course recognizes that individual causes might, in rare circumstances, suffice for their effects; and that a set of causes might in rare circumstances be necessary for its effect. Hence he proceeds to coin the rather ugly phrase "at least an INUS condition". But this is all a minor terminological blunder. The sensible thing to do is to proceed as Bennett (1988) does, and take causes to be necessary parts of sufficient conditions for their effects, or "NS conditions" for short.

§2.3 "Conditions" vs. events

[1] Bennett is quite good on this point; see his 1988.

What exactly is a "condition"? Far from picking out a familiar ontological category, the word seems mainly useful as a sort of dummy term, to be employed when one is not exactly clear about the ontology of the causal relation. Inspecting Mackie's writings, though, one quickly sees that he is quite happy to subject "conditions" to all the usual truth functional operations—negation, disjunction, conjunction, etc.—a feature of his account that suggests that he has in mind *facts* or perhaps *states of affairs* as the causal relata.

That choice is a serious mistake, and we need not clarify the rest of Mackie's account (e.g., the notion of "sufficiency") to see why. Consider, for example, the following situation (figure 4.1):

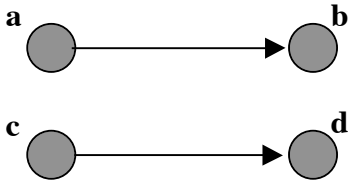

Figure 4.1

Here, **a** fires, sending a stimulatory signal to **b**, causing it to fire. Simultaneously, **c** fires, sending a stimulatory signal to **d**, causing it to fire. Let A be the fact that **a** fires (if you like, the "condition" of **a**'s firing), B the fact that **b** fires, etc. We would like to say that A causes B, but that *c* does not. Presumably, then, A is in some sense sufficient for B—but, in whatever sense of "sufficient" that is, C is not. Let us suppose that that sense is roughly the following: A set of conditions is sufficient for some effect just in case (i) those conditions in fact obtain; and (ii) the effect follows from the conditions, together with the laws, together perhaps with some suitably chosen auxiliary proposition. In the present case: B follows from A, together with the laws, together with some suitably chosen auxiliary proposition describing the circumstances in which **a** fires (perhaps that neuron **a** is connected in the right way to neuron **b**, and that there are no interfering factors in the environment—more on this shortly). But then, in exactly the same sense, the pair {C, either not-C or A} will be sufficient for B. What's more, C is an essential element of this set, in the sense that if we remove it, the resulting set {either not-C or A} no

longer suffices for B. There appears to be no hope of fixing this problem merely by tinkering with the account of sufficiency.

This is not a petty objection. Rather, it shows at some quite fundamental level that it is wrongheaded to pursue a Mackie-style analysis of causation while taking facts—or some other ontological kind "closed" under truth-functional operations—to be the causal relata.[2] The natural response? Pick a kind not so closed: namely, *events*.

Of course, one may attempt to stick with facts by trying to impose restrictions on which facts are allowed to stand in causal relations. No doubt that thought will have already crossed your mind: Isn't there something unduly *disjunctive* about the fact (either not-C or A) that should disqualify it at the start as a candidate cause of B? Perhaps—but this something cannot be merely that we refer to this fact by means of a disjunction. In order for it to be a feature of the *fact* that it is disjunctive, and not merely a feature of our means of expressing it, we will presumably need something like an objective division of the facts into the atomic and the non-atomic. (Maybe some other distinction, similar in spirit, would do.) And indeed, the events depicted in figure 4.1 seem to provide such grounds, as witness my natural reference to them as "events". That is, we might say that A counts as atomic because it corresponds to *a*—an *event*, the firing of **a** at a particular time. Likewise C corresponds to *c*, the firing of **c**. But the "condition" (either not-C or A) does not correspond to any event; there simply is no event which somehow consists in either the non-firing of **c** or the firing of **a**. So perhaps we should say, in general, that the "atomic" facts—or at any rate, the facts suitable to stand in causal relations—are those facts to the effect that some particular event occurs.

So far, that maneuver looks suspiciously like a concession that, after all, *events* are the fundamental causal relata. So perhaps we are back to what I called the "natural" response. Then again, perhaps not: for there remains one reason for sticking with facts that needs to be

[2] Curiously, Bennett seems not to have noticed this problem. In *Events and Their Names*, his preferred analysis of causation is exactly a Mackie style analysis taking facts as the relata—and it runs straight into the problem we have just raised. For a much more detailed discussion of the problem, see Kim 1971.

addressed, although doing so properly will require a chapter unto itself (chapter 9). Specifically, one might wish to allow that there is such a thing is causation by omission (and prevention—which perhaps should be thought of as causation *of* omission); but one might hope that doing so will not require adding to one's ontology such monstrosities as "negative events". There are plenty of reasons to shun them; we will review these in chapter 9. These reasons seem all the more compelling, given how easy it appears to be for a certain kind of *fact* to take the place of "negative events". When, for example, we assert that the failure of the alarm to sound caused the house to burn down, we can take ourselves to mean not that the house's burning down was caused by a mysterious sort of event—one that "occurs" just in case the alarm does not sound—but rather by the fact that the alarm did not sound (some will prefer: the state of affairs that consisted in the alarm not sounding). More generally, causation by omission statements can be taken to have the canonical form: The fact that no event of type C occurred caused *e*.[3] Finally, admitting facts such as these as candidate causal relata need not force us to take the further step of admitting the troublesome "disjunctive" facts discussed above.

I think this would be a good reason for taking fact causation seriously[4], were there not overriding reasons to give a quite different treatment of causation involving "omissions" from the treatment appropriate for ordinary causation. In chapter 9, in fact, I will argue that an adequate philosophical account of causation need not and indeed should not treat ordinary kinds of causation, on the one hand, and causation by omission and prevention, on the other, in a uniform manner; in a certain sense, these latter kinds should be treated as special cases. This chapter will lay some of the groundwork, for it will emerge, in our development of a regularity analysis of causation, that there is no straightforward way to handle causation by omission or prevention. There are principled reasons for this shortcoming—principled in the sense that they suggest that the fault lies not with the regularity analysis, but with the thought that these kinds of

[3] For reasons to be discussed in chapter 9, that form seems better to me than the following alternative: The fact that event *c* does not occur causes *e*.

[4] See Mellor 2003 for an example of a philosopher who endorses this reason without qualification.

causation should be treated just like ordinary causation. As indicated, we will see some of these reasons in this chapter; but the next two chapters, as well as chapter 9, will add important new reasons.

So I conclude that we should take the causal relata to be events. Indeed, many philosophers of causation would have started out with this position, treating it as more or less intuitively obvious. As already noted (chapter 2), I don't think it is particularly intuitively obvious, nor that the usual arguments against the rival "fact causation" view are much good. Hence it has seemed worthwhile to point out why, given that we are pursuing a regularity analysis that tries to build on the main idea behind Mackie's own, the choice of events as the primary cause relata does after all carry important advantages.

In making this choice, we are taking several strides away from Mackie's own treatment of causation, since we no longer have available to us all the luxuries of conjunction, disjunction, negation, etc. (to borrow a phrase from Bennett). As we'll see, that's no loss.

§2.4 Restricting candidate sets of causes

We have reached the following point: A cause of an effect is an event belonging to a set of events that suffices for that effect, and that contains no non-essential elements. Call a set S *minimally sufficient* for event *e* just in case S is sufficient for *e*, but no proper subset of S is sufficient for *e*. Then we could also say: A cause of an event *e* is a member of a set of events minimally sufficient for *e*. The next problem to solve comes into focus when we see that on any reasonable account of sufficiency, a cause will belong to too many sets minimally sufficient for its effect; we must therefore restrict which events are allowed to form such sets.

(a) Temporal precedence

This problem has two aspects. For the first, observe that an event may well have *effects* that suffice for it. As a simple example, consider figure 4.1 again. We would like our account of

sufficiency to secure the verdict that the set {*a*} is sufficient for *b*.[5] But how are we to construct an account of sufficiency that gets *this* right without *also* yielding the conclusion that {*b*} is sufficient for *a*? We face, in other words, the problem of designing an account of causation that will guarantee that that relation is (typically, at least) *asymmetric*. I propose to solve this problem in perhaps the crudest possible matter, by requiring that causes *precede* their effects. The costs are obvious: We rule out, a priori, the possibility of both backwards and simultaneous causation. What is much less obvious than meets the eye is exactly how steep those costs are. In chapter 12, I will argue that they have typically been wildly overestimated. In some cases, the possibilities prove illusory; in others, we can find well motivated ways to accommodate them as special cases. What's more, the popular claim that counterfactual analyses can easily accommodate backwards and simultaneous causation proves, on inspection, to be deeply mistaken.

The reader may have been struck at this point by a different, and apparently much more devastating objection—more devastating, because it doesn't need any help from outré possibilities whose legitimacy might be denied. Consider the following diagram:

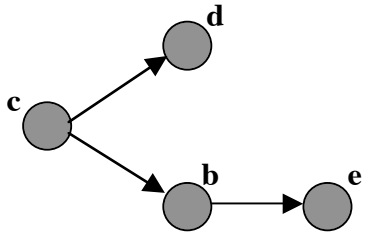

Figure 4.2

Here **c** fires, causing **d** to fire; along another pathway, **c**'s firing causes **e** to fire. *d* occurs before *e*. It might also seem, given the way the neuron network is set up, that {*d*} is sufficient for *e*: The thought is, roughly, that from the fact that **d** fires, together with the laws, together with suitable auxiliary premises describing how the neuron network is set up, one can validly infer that **e** fired (because **c** must have fired in order for **d** to fire, and therefore **b** must have fired as a result of **c**'s

[5] No doubt this set should contain other events as well—e.g., the events consisting in the prior presence of the stimulatory channel connecting **a** to **b**, and the prior presence of neuron **b** itself. For ease of exposition, I'm going to ignore these other events.

firing, etc.). If so, then our sufficient condition disastrously conflates causes with joint effects of a common cause.

As a response to the maneuver of guaranteeing asymmetry by requiring that causes precede their effects, this objection looks quite nifty—so nifty that it is not very surprising that philosophers have failed to see that it doesn't succeed. We will turn in §3 to the details; for now, I will simply foreshadow: The best and most obvious way to construct an account of sufficiency simply does not fall prey to this objection. In figure 4.2, {*d*} is not, after all, sufficient for *e*.

(b) Causal combination, and uniqueness

But there is a second aspect to the problem under discussion not at all touched by our stipulation about temporal priority. Consider a case, as in figure 4.3, where multiple causes act jointly to bring about a given effect:

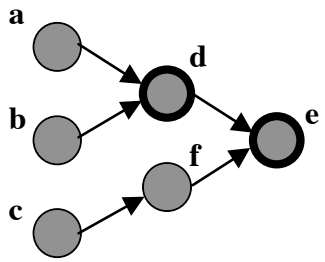

Figure 4.3

Here, neurons **d** and **e** both require two stimulatory signals as input in order to fire, a fact represented by drawing them with thick borders. Thus, the firings of **a**, **b**, and **c** act jointly to bring about the firing of **e**, partly by way of the firing of **d**. Now focus on *c*. It is a cause of *e*. But with what other events (that precede *e*—we will mostly omit this qualification henceforth) should it be combined in order to form a set minimally sufficient for *e*? There are, unfortunately, an embarrassingly large number of choices. Consider just two: the set {*a*, *b*, *c*} and the set {*c*, *d*}. It might not yet be clear what the problem is; why not count *c* a cause of *e* just in case it belongs to it least *one* set of events minimally sufficient for *e*?

But simple examples of overdetermination, such as that depicted in figure 4.4, decisively rule this option out:

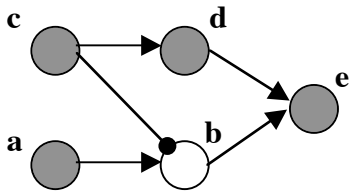

Figure 4.4

Here, the set {*a*} is minimally sufficient for *e*; but *a* is obviously not a cause of *e*. It might be hoped that we could avoid this consequence by being very careful about how we define "sufficient": maybe there is a definition that will render {*c*} sufficient but {*a*} not. I think there is no future in this option. Rather, I think we should draw two lessons from the current discussion: The first is that the test for causation needs to be much more stringent than the one just suggested. It will not do to say that *c* is a cause of *e* iff it belongs to *some* set of events minimally sufficient for *e*. The simplest fix, which I endorse, is to require that *c* belong to a *unique* set minimally sufficient for *e*. (Later, in §4, we will see the need for some amendments; they do not matter, for present purposes.) Having done so, we cannot of course claim to be providing a *necessary* condition on causation (else overdetermined events will come out uncaused), and so we will eventually face the problem of how to complete our analysis. But for now, let us take it that we are merely trying to provide a *sufficient* condition for causation.

There is a second lesson to incorporate. For unless we place substantive restrictions on which events can be gathered together into minimally sufficient sets, the requirement of uniqueness will almost never be satisfied: In any case where the target effect is brought about by way of multiple converging causal pathways we will be able to find, for most if not all of its causes, multiple minimally sufficient sets for them to belong to. We need merely choose, as events to accompany a given cause, any of the many causes belonging to an independent pathway. That is exactly what happened in our discussion of figure 4.3: we could combine *c* either with the contemporaneous events *a* and *b*, or with the slightly later event *d*. Each of these choices is guaranteed to work, given the position of the other events in causal pathways independent of the one to which *c* belongs. (Notice that combining *c* with *f* poses no problem: In general, combining

events from the same causal pathway will, as here, simply produce a set that is sufficient, but not minimally so.)

If we go back to the guiding idea, I think we can see a principled way to impose the needed restriction. When we have the thought that what is distinctive about causes is that they *combine together* to form sets minimally sufficient for their effects, the notion of "combine" we have in mind is not a stripped-down set-theoretic one; it is, rather, a *causal* one. That is, we think that when an event *c* causes an event *e*, *c* will typically *act together* with other events to bring about *e*. So the straightforward way to restrict our choice of what events can form a minimally sufficient set is to require that these events "act together". Returning to figure 4.3, we see a distinction: Intuitively, *c* acts together with *a* and *b* to bring about *e*; *c* does not act together with *d*—rather, it is *f* that acts together with *d* to bring about *e*.

This notion of "acting together", or, as I will sometimes put it, "causal combination", is one we will have to return to later on (chapters 6 and 12). For the present, it is best to replace it by a surrogate notion that is more precise, and whose use won't introduce unanalyzed causal notions into our analysis of causation. So I will simply stipulate that minimally sufficient sets must be formed out of *contemporaneous* events—taking my cue here from the fact that in figure 4.3, the events that causally combine to bring about *e* are contemporaneous.

(c) Causation across temporal gaps, and uncaused events

This move has consequences that we should pause over, making trouble for accommodating certain kinds of abnormal causation. First, there is action at a temporal distance. Here is a fanciful example:

Figure 4.5

Imagine, in figure 4.5, that **c** is a special sort of neuron which, when it fires, acts at a temporal distance on **e**. That is, it is a law that whenever **c** fires, then a certain specified amount of time later, **e** fires. We represent this fact by drawing a fat, faded out arrow between **c** and **e**; note well

that this arrow does not depict a stimulatory channel, nor the passage of a signal along such a channel. In fact, we can suppose that *c* leaves no physical trace of itself *whatsoever* in the physical states of the world between the time of its occurrence and the time at which *e* fires. (Not an essential stipulation—but it helps the imagination.)

Now make the situation a bit more complicated, as in figure 4.6:

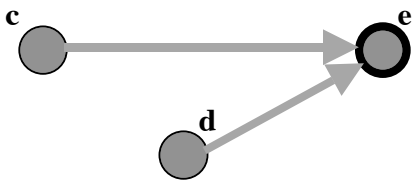

Figure 4.6

This time, **e** needs *two* stimulatory "signals" acting at a temporal distance in order to fire. It gets one from **c**, which fires at time 0. It gets the other from **d**, which fires at time 1. And so **e** itself fires at time 2. As before, we can suppose that no other events *whatsoever* occur, besides these three. Neither *c* nor *d* causes *e* all by itself. Rather, these events *jointly* bring about *e*. But there is no way to choose *contemporaneous* events that include *c* or *d*, and that will form a set sufficient for *e*: that is because there are no events *at all* contemporaneous with *c* and *d*, but neither *c* nor *d* is sufficient for *e* all by itself. We thus face a serious problem if we wish to accommodate this kind of action at a temporal distance.

Exactly the same problem arises if we allow, as a genuine possibility, the spontaneous, uncaused occurrence of events. Of course, if we think that time has a first moment, we have *already* allowed this possibility, since the events occurring at that first moment presumably lack causes. But I have in mind something much more bizarre, which is that after the first moment of time (if such there be), events sometimes spontaneously occur, without any prior causes. So consider another fanciful scenario:

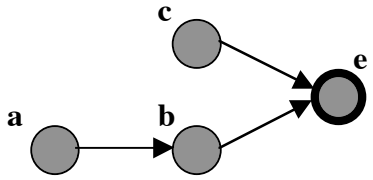

Figure 4.7

Here, we can suppose that **a** fires as a result of ordinary processes. A short time later, its stimulatory signal causes **b** to fire. Meanwhile, and simultaneously, **c** fires *spontaneously*. The signals from **c** and **b** converge on **e**, causing it to fire. Let us finally add that **e** requires both stimulatory signals in order to fire. Clearly, *a* is a cause of *e*. But once again, we can suppose that there are no events contemporaneous with it at all, hence no contemporaneous events with which it can form a set minimally sufficient for *e*. Moreover, that **e** needs two stimulatory signals in order to fire guarantees that {*a*} is not sufficient for *e*.

One might expect, at this point, a reminder that we were only intending to put forth a *sufficient* condition for causation. Hence, the examples just given do not threaten an outright contradiction. Fair enough. But recall why exactly it was that we limited ourselves, for the moment, to a sufficient condition: we wanted to avoid the trouble that would otherwise be generated by simple cases of overdetermination. Later (§4, below, and chapter 6), we will see how to develop techniques that are both intuitive and extremely well motivated for extending our sufficient condition so as to allow us to handle overdetermination. We will see that those techniques may help us handle spontaneous causation; but they will allow us to make no progress at all with action at a temporal distance. Nor will it help to go back to the drawing board, and try to come up with a replacement for the requirement that events forming a minimally sufficient set be contemporaneous. For that replacement would presumably have to pick out, in figure 4.6, a unique event for *d* to combine with. Should this event be *c*? Why not one of *c*'s own causal precursors, instead? There seems to be no non-arbitrary way to make this choice—but a choice must be made, for we need to find a *unique* minimally sufficient set for *d* to belong to. Moreover, even if we settle for an arbitrary way of making this choice, how are we to extend this way to ordinary cases involving no action at a temporal distance? In cases of action at a temporal distance we want to allow events to form minimally sufficient sets with events occurring at different times; on what basis do we forbid this, in the ordinary case?

High time, I think, to beat a hasty retreat, and to resolve to take up the thorny issues involving abnormal causation at a later point (chapter 12). Happily, we will find both ample grounds and ample means for treating the kinds of examples discussed here as special cases.

§2.5 Defining "sufficient"

Let us take stock. We have arrived at the following (provisional) sufficient condition for causation: Event *c* is a cause of later event *e* if it belongs to a set of contemporaneous events that is minimally sufficient for *e*, and that is unique in this regard. Now we need to explain what "sufficient" means. Recall that we are assuming that the fundamental laws are deterministic. A good thing, too: for if they are indeterministic, no reasonable account of sufficiency will get the result that the causes of an event occurring at a given time are sufficient for it. (Just consider that they need not even bestow upon it a particularly high probability.) Causation under indeterminism is a hard enough topic to deserve its own chapter; for now, rest assured that when we get there (chapter 13), the main ideas being developed for the simpler case of causation under determinism will have a natural extension.

An obvious but naïve thought is that by restricting ourselves to determinism, we have made easy work of defining "sufficient": A set of events S is sufficient for *e* just in case, from the fact that every event in S occurs, together with the laws, it follows that *e* occurs. But there are three problems with this definition, ranging from mild to devastating. Let us take them up in that order.

(a) The notion of entailment

First, we need to specify the sense of "follows". Do we mean that from a sentence expressing the fundamental laws, together with a sentence, for each event in the given set, expressing the proposition that that event occurs, one is able to deduce (in accordance with the rules of some appropriate formal deductive system) a sentence expressing the proposition that the effect occurs? Certainly not—to point out just one of the most obvious problems, the sentence expressing the laws will of course be bereft of any proper names for events, so there is no hope

of deducing a conclusion of the form "*e* occurs", where "*e*" is a name for the effect. (Readers familiar with Davidson may wonder whether it will help to make use of sentences that *describe* each event—presumably, in the language of fundamental physics. For the reasons discussed in chapter 2, I think this Davidsonian approach is a poor idea.)

Better to understand "follows" as "metaphysically necessitates". Then we can state the definition a bit more carefully: S is sufficient for *e* just in case there is no possible world in which the actual laws hold, and in which every member of S occurs, but in which *e* does not occur. So far, so good.

(b) Essences of events

But two problems remain, and they are harder. The first problem is that it is not at all clear what worlds we have in mind, and that is because it is not at all clear, for any given event, what marks the boundary between those worlds in which it occurs and those in which it does not. The literature tends to overlook this problem (Bennett 1988 is a notable and welcome exception). For example, in his classic "Causation"—the very paper that launched the thriving tradition of attempting to give a counterfactual analysis of causation—Lewis writes, without blinking,

> We have spoken thus far of counterfactual dependence among propositions, not among events. Whatever particular events may be, presumably they are not propositions. But that is no problem, since they can at least be paired with propositions. To any possible event *e*, there corresponds the proposition O(*e*) that holds at all and only those worlds where *e* occurs. This O(*e*) is the proposition that *e* occurs. (1973, p. 166)

One might wonder what exactly this proposition is, and a hint that it will not be easy to find out is provided by the very long footnote Lewis appends to the passage just quoted. This footnote is worth reproducing in full:

> Beware: if we refer to a particular event *e* by means of some description that *e* satisfies, then we must take care not to confuse O(e), the proposition that *e* itself occurs, with the different proposition that some event or other occurs which satisfies the description. It is a contingent matter, in general, what events satisfy what descriptions. Let *e* be the death of Socrates—the death he actually died, to be distinguished from all the different deaths he might have died instead. Suppose that Socrates had fled, only to be eaten by a lion. Then *e* would not have occurred, and O(e) would have been false; but a different event would have satisfied the description "the death of Socrates" that I used to refer to *e*. Or

> suppose that Socrates had lived and died just as he actually did, and afterwards was resurrected and killed again and resurrected again, and finally became immortal. Then no event would have satisfied the description. (Even if the temporary deaths are real deaths, neither of the two can be *the* death.) But *e* would have occurred, and O(*e*) would have been true. Call a description of an event *e rigid* iff (1) nothing but *e* could possibly satisfy it, and (2) *e* could not possibly occur without satisfying it. I have claimed that even such commonplace descriptions as "the death of Socrates" are nonrigid, and in fact I think that rigid descriptions of events are hard to find. That would be a problem for anyone who needed to associate with every possible event *e* a sentence Φ(*e*) true at all and only those worlds where *e* occurs. But we need no such sentences—only propositions, which may or may not have expressions in our language. (*ibid.*)

Let us agree with Lewis, as we should, that even when we refer to an event by means of a uniquely identifying description containing no demonstrative elements, that description almost never captures the exact conditions of occurrence for the event. That is, we cannot simply appeal to that description when trying to say what marks the difference between worlds in which the event occurs and worlds in which it doesn't. The problem is that once we have agreed to that much, we are in no position to adopt the blase attitude he manifests in this footnote. Observe, in this regard, how odd the closing sentences are—and how they betray the papering over of a serious issue. Lewis writes, "That would be a problem for anyone who needed to associate with every possible event *e* a sentence Φ(*e*) true at all and only those worlds where *e* occurs. But we need no such sentences—only propositions, which may or may not have expressions in our language." But of course by his lights, such sentences are *ridiculously* easy to come by: as witness his example of "*e*", which he uses as a rigid name for the death of Socrates (a use not to be confused with his earlier and later employment of this letter as a variable ranging over events). Thus, for any event *e*, all we need to do is to introduce into our language a name for that event, stipulating (if necessary) that this name is to refer rigidly. Suppose, for example, that you and I have a conversation, and we decide after the fact to call that conversation "Fred". Then we have our sentence, which is true at all and only those worlds where Fred occurs: it is just "Fred occurs". Why has Lewis missed this trivial point?

Of course I don't know, but I have some conjectures. Given any thing X—be it an event or ordinary object—talk of the proposition that X exists or occurs can easily strike us as unproblematic (bracketing nominalist worries, etc.). However, as soon as we identify such

propositions with sets of possible worlds, trouble arises, in the form of questions that we seem unable to answer concerning the distinction between those worlds in which X does and those in which it does not exist or occur. Under what conditions, for example, would this very computer (the one in front of me) have existed? Could it have been a work of post-modern art, if its constituent parts had been arranged differently? Or is it essential to it that it be a computer? A working computer? Who knows? Confronted with questions like these, we often throw up our hands in a way symptomatic of the absence of any determinate answer.

With ordinary objects such indeterminacy has its limits, as witness the ease with which we can often reason about what would have happened to a given thing under such-and-such counterfactual circumstances. But matters appear otherwise with events, where the indeterminacy seems much more acute. Take Fred: Would it have existed—would we have had numerically the same conversation—if instead of talking about philosophy we had talked about the weather? Did it need to be a conversation at all—could Fred instead have been a duet that we were singing? Could it have taken place on a different day? Could some other pair of people have had it? Could it have had substantially different causes and effects? I think that for all of these questions—and, obviously, for an untold multitude of others—there is absolutely no principled basis for a precise answer.

There are natural constraints: Arguably, Fred could not have been a nap that one of us was taking, etc. Arguably, Fred still would have occurred in a world that differed only in respects *extrinsic* to its spacetime region. But such constraints don't take us very far in the direction of a proper account of the occurrence-conditions of events. For example, it remains unclear whether even the merest difference in manner of occurrence would result in a numerically different event. And that means that we have no business helping ourselves to the assumption that, for any given event, there is even a reasonably clear distinction to be drawn between those worlds in which the event does and those in which it does not occur. Rather, we have serious work to do: the work of coming up with an account of this distinction (a stipulative one will do!), one that will properly serve the needs of our account of causation.

What's more, unless we keep Lewis's strictures firmly in mind, it is very easy miss the need for such an account by sliding between locutions of the form "the Φ–□□□ of X at t occurs" to more familiar locutions of the form "X Φ□□ at t". For example, this slide makes it vastly easier to evaluate the counterfactuals deployed in typical counterfactual analyses. Suppose that neuron **a** fires at noon. To investigate its effects, the counterfactual analyst starts by asking what would have happened if *that very firing* had not occurred. A strange question! (Are we supposed to consider scenarios in which **a** fires a little later or a little earlier, or perhaps in a somewhat different manner—taking these to be situations in which numerically the same firing does not occur? Who knows—and who on earth knows how to find out?) How much easier it is to ask what would have happened if, at noon, **a** had not fired: for it will almost always be perfectly clear how to construct a counterfactual scenario that satisfies *this* antecedent. In short, then, I think that undue charity towards counterfactual analyses leads us almost automatically to replace the *official*—and very difficult to understand—counterfactuals used in such analyses with *unofficial* ones that are much easier to understand.

Returning to our diagnosis of Lewis, let me just observe that it is exactly these issues that come into sharp relief when we introduce a name for an event, and understand it to be a rigid designator—for to say that it is a rigid designator is just to say that it refers, in any possible world, to that very event (if the event exists).

Now, the issue we have raised here is a subtle and difficult one; in the face of its subtlety and difficulty, I am going to follow a usefully craven policy, and pretend for the moment that it doesn't exist. But just for the moment: We will take up the question of event "essences" again shortly, and provide a much more extensive treatment in chapter 7. It should also be emphasized that as long as we are sticking to sanitized examples involving neuron diagrams, we will mostly be able to stay out of trouble. For in such cases, it is clear enough—or at least, clear enough for most purposes—where the line should be drawn between worlds in which a given neuron firing occurs and worlds in which it doesn't.

Remember at this point that that was just the second of three problems that the proposed definition of sufficiency raised. Let us turn now to the third.

(c) Background conditions, and the "no-interference" clause

The third problem for the proposed definition of sufficiency—embarrassingly obvious, once pointed out—is that pretty much nothing ever satisfies it.

Some events occur at time t. They happen to be all the causes occurring at that time of some later event *e*. Does it follow from the fact that all of these causes occur, together with the laws, that *e* occurs? Of course not. For it is consistent to add, as an extra premise, that still other events occur—events that initiate processes that get in the way of the causes, preventing them from bringing about *e*. (Nothing in fact interferes, but that is beside the point.) In order to salvage our definition, it looks like we need to add some sort of "no-interference" clause.

Distinguish this problem from the problem that I suspect most philosophers think bedevils the regularity analyst's attempt to come up with an adequate account of sufficiency. I will rehearse this other problem in the way it was first taught to me.

Begin with a regularity analysis that takes the following form: *c* causes *e* iff from the fact that *c* occurs, together with the laws, together with suitably chosen true propositions of particular fact (detailing the circumstances in which *c* occurs), it follows that *e* occurs. To review, we have already amended this analysis in significant respects: we have replaced the "iff" with an "if", and we have focused not on single events but on sets of contemporaneous events. Still, the key idea that we need some suitably chosen true proposition describing the *background conditions* in which the putative causes all occur seems to lead to serious trouble. Consider figure 4.2 again.

Here, we want it to come out that *c* is a cause of *e*. In order to do so, we need to include as a premise a true proposition describing the way that *c* is connected up to *e*. Suppose that Φ is this proposition, delivered by whatever principles we use in selecting such auxiliary premises. So we have the following attempt at an analysis: from Φ, together with the fact that *c* occurs, together with the laws, it follows that *e* occurs. But it had better *not* be the case that from Φ, together with the fact that *d* occurs, together with the laws, it follows that *e* occurs. Now comes the problem: It

is just very hard to see what principled way for choosing the needed auxiliary premise Φ could guarantee that the first entailment will go through but not the second. This particular problem has come to be known in the literature as the "problem of effects"—so called, because analyses that fall prey to it do so by failing to distinguish genuine *causes* from *joint effects* of a common cause.

As stated, the problem runs together two separate issues by conflating two distinct roles that the auxiliary premise is supposed to play. On the one hand, we need, as we have seen, an auxiliary premise to the effect that *nothing interferes* with the putative causes. On the other hand, we need an auxiliary premise specifying those *background facts* in virtue of which the causes are able to bring about their effects. As an illustration, in figure 4.2 we need our auxiliary premise to say that neuron **c** is connected by stimulatory channels and intermediate neuron **b** to neuron **e** (that is the second role); but we also need it to say that there are no inhibitory signals or other potentially interfering processes threatening to prevent the stimulatory signal from reaching **e** (that is the first role).

It is very important to keep these two roles sharply distinguished. The reasons why will be more or less implicit in the discussion that follows.

Let me begin by commenting on the second role. Arguably, nothing need play it—not, at least, if we are willing to take a liberal enough view about what counts as an event. In analyzing figure 4.2, for example, we started by letting our set of candidate causes for *e* (occurring at time t, the time of **c**'s firing) include just *c* itself. But if we count boring non-changes as events, as I suggested in chapter 2, then we should really include in this set events consisting in the presence at time t of neurons **b** and **e**, as well as events consisting in the presence at time t of the stimulatory channels connecting **c** to **e** via **b**. In other words, we carve up what are usually thought of as "the background circumstances" into a number of *events*.

I do not mean to insist on such a liberal construal of events. In fact, I think it can be quite important to attend to the distinction between events and background circumstances, particularly if one is interested in providing an accurate map of our ordinary causal concepts (as opposed to

providing tidied up partial revisions of them). All the same, it is worth noting how this way of setting things up allows us to frame the problem of effects much more sharply. For it now appears easy to construct a set of contemporaneous events minimally sufficient for *e*, and that contains *d*. We simply let the set contain, in addition to *d*, the following "events": first, the presence, at the time of **d's** firing, of the stimulatory channel connecting **c** with **d**; second, the presence at that time of the stimulatory channel connecting **c** with **b**; third, the presence of the channel between **b** and **e**; fourth, fifth, and sixth, the presence of the neurons **b**, **c**, and **e**. Note that we do not include either the firing of **c** or the firing of **b**, for that would make *d* redundant, and the set no longer minimally sufficient for *e*. As it stands, though, it apparently *is* minimally sufficient for *e*: the occurrence of all these events seems to guarantee (modulo the provision of an appropriate no-interference clause, and together with the laws) that *e* occurs (by guaranteeing that **c** must have fired, etc.); but no proper subset would do so. (Shortly, we will see that this seeming guarantee is in fact merely that.)

For now, I will not insist that we redescribe background circumstances as being composed of events, and will take it that one auxiliary premise (the one playing the second role) describes all relevant background circumstances. More specifically, in the examples to come we can take this auxiliary premise to completely describe the neuron network being depicted. Then our problem is that given that the neuron network is set up as described, and given that there are no interfering factors, and given the laws, it seems that if *d* occurs, then *e* must occur. For it seems that if *d* occurs, then it must have been that *c* occurred, etc.

Let's set this problem aside for the moment, and take up the question of how to provide, in a principled way, a suitable no-interference premise with which to round out our definition of sufficiency. At first sight, this might seem an impossible task. Suppose, for example, that we go about things in as flat-footed a manner as possible, saying that a set of contemporaneous events S is sufficient for some later event *e* just in case it follows from the fact that the events in S occur, together with the laws, together with the additional premise that nothing *interferes* with the events in S (or perhaps better: with the processes these events initiate), that *e* occurs. There are

two problems. First, since "interferes" clearly labels some kind of *causal* relation, we have given up rather too quickly on the aim of providing a reductive analysis of causation. Second, the premise is too imprecise, since what we really want is something like this: Nothing interferes with the events in S *in their attempt to bring about e*. The point is that the events in S might be, as it were, apt for bringing about other events besides *e*—but might, because of external circumstances, be prevented from doing so. Consider for example the following diagram (figure 4.8):

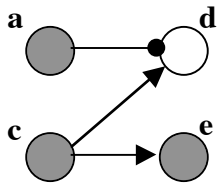

Figure 4.8

Here, **c** fires, sending stimulatory signals to both **d** and **e**. But **a** fires at the same time, sending an inhibitory signal to **d**; hence while **e** fires, **d** does not.

Suppose we try to say that {*c*} is sufficient for *e* because it follows from the fact that **c** fires, together with the laws, together with a proposition describing the given neuron network, and together finally with the premise that nothing interferes with *c*, that **e** fires. Then it would seem that this last no-interference premise is *false*, since *a* does in fact interfere with *c* in its attempt to bring about the firing of **d**. Of course that wasn't what was intended; what we had in mind was that nothing interferes with *c* with respect to the production of *e*. Alas, once we say this explicitly, the analysis appears to lose whatever interest it may have had. Finally, notice that in figure 4.2 nothing at all interferes with the firing of **d**; so the problem of effects is apparently still very much with us.

Let's rewind. It turns out that at the outset, we unblinkingly accepted a claim that should have aroused at least mild suspicion. Remember the dialectical sequence: We started by observing that from the fact that some contemporaneous events occur, together with the laws, it need not follow that some given effect of those events occurs; we must add, as additional premises, a true proposition describing the background circumstances, and another true

proposition that somehow "says" that nothing interferes in the generation of the effect by those causes. *True* propositions, note. But why was that? Why did these auxiliary premises have to be true? Well, it's perhaps clear enough in the case of the auxiliary premise that details the background circumstances; after all, *falsifying* the background circumstances could all too easily generate spurious connections between non-causes and some given effect. But why did the no-interference premise need to be true? On reflection, the answer is not at all clear. More: it *is* clear that it did *not* need to be true. But to see why this is so, we should back way up, and come at the issues we have been dealing with from a somewhat different direction.

Suppose that at a certain time t, the events in set S are the *only* events that are occurring. A little while later, event *e* occurs. Then isn't it clear that the causes of *e* at t are to be found within the set S (and that, moreover, there *are* causes of *e* in this set)? "Well of course," comes the reply, "after all, *e* doesn't occur spontaneously, so where *else* could its t-causes be found?" But now let us add something more. Suppose in addition that for any proper subset S' of S, in a nomologically copossible situation where only the events in S' occur at t, *e* does *not* occur. Then it seems that we can say something stronger about the original situation: not only are the t-causes of *e* to be found in S, but *all* the elements of S are t-causes of *e*. (That's not quite right, for reasons that will emerge later, and that don't matter now.) Observe that in these circumstances, it is quite easy to provide no-interference clauses for the sets S and S'; the needed premise is simply the claim that at t, no other events are occurring. Then why not adopt this sort of claim in general as our no-interference premise? We should say, that is, that a set S of contemporaneous events is sufficient for some later event *e* just in case it follows from the fact that every event in S occurs, together with the laws, together with a proposition describing background circumstances, together finally with a proposition that nothing *else* occurs at the time of occurrence of the events in S, that *e* occurs.

Objection: But this last premise will, in typical cases, be *false*.

Reply: So what? There is, in the first place, no *logical* obstacle to letting this proposition be false—this makes no difference, after all, to the argument's *validity*. And, much more

importantly, there is an attractive way of thinking about sufficiency that makes this proposition seem an obvious choice for the no-interference premise. The idea is this: When we say that a set of contemporaneous events is sufficient for some later event, we have in mind that these events are *apt for producing* that later event—they are, if you like, *disposed* to be followed by it, and will manifest this disposition provided nothing gets in the way. But to think of sufficiency for an event *e* as a kind of dispositional property of a set of events—one that can be masked or mimicked by external circumstances—immediately suggests a test for whether it is present: simply place the events that purport to possess it into circumstances where there is no threat of masking or mimicking. The simplest choice? Circumstances in which nothing else is happening. Compare: In order to determine whether some substance is disposed to dissolve in water, don't simply ask whether it would dissolve, were it put in water; for there might be factors in the environment that would either mask the disposition it in fact possesses, or mimic the disposition it in fact fails to possess (see Lewis 1997). Rather, ask whether, if it were put in some water in circumstances where nothing else existed, it would dissolve. That way, the confounding presence of maskers or mimickers ceases to be a concern. In short, sufficiency for *e* is aptness to be followed by *e*; a straightforward test for the presence of this disposition is to place the events in question into circumstances in which they are, as it were, *left to themselves*.

A slight refinement is in order. Thus far, our sufficient condition asks us to look to nomologically possible worlds in which the events in the given set S are occurring at the given time t, but nothing else is occurring at that time. One problem with this test is that it requires us to distinguish worlds in which it is the events in S that are occurring from worlds in which it is events very similar to, but numerically distinct from the events in S that are occurring. We can finesse this problem—and at the same time render our account of sufficiency much more intuitive—by stating that account in counterfactual terms, thus: say that a set of events S occurring at time t is sufficient for later event *e* just in case, were the events in S the only events

to be occurring at time t, *e* would still occur.[6] That still requires us to consider a counterfactual situation in which it is the events in S that are occurring, but we can arrive at this counterfactual t-state by beginning with the actual state of the world at time t, and making minimal modifications sufficient to guarantee that nothing besides the events in S is occurring at time t. (Nothing, *anywhere?* Yep. It's a strange counterfactual situation—very different from the actual world—but that matters not at all to the uses to which I put it.) Since these modifications, however drastic they may be, will concern factors extrinsic to the events in S, we can safely assume that the counterfactual situation thus arrived at is one in which it is the events in S that are occurring, and not "doppelganger" events very similar to but numerically distinct from these events.

Finally, a note about background conditions: If we treat them as just so many events, then no modification to our account of sufficiency is called for. On the other hand, if we wish to adhere to a distinction between genuine events and mere background conditions, then we should understand that in the counterfactual situation in which only the events in S are occurring, the actual background conditions are to be held fixed.

(d) Fine-grained causation

There are two problems that I must simply flag for future discussion (chapter 7). First, this counterfactual test for sufficiency explicitly ignores, as irrelevant to the causal status of the events in S, nomologically possible situations in which only the events in S are occurring at t, but one or more of these events is occurring in a manner substantially *different* from the manner in which it actually occurs. Many authors would claim that we ignore such situations only at our peril. A simple example will illustrate why. Suppose, in figure 4.9, that **c** fires intensely (hence its dark shading):

---

[6] It has been suggested in conversation that this appeal to a counterfactual makes my analysis a counterfactual analysis. I don't know whether that is supposed to be an objection. If so, it is obviously silly. At any rate, I don't care about the name, as long as its use does not obscure the fundamental differences between my account and any account that puts counterfactual dependence between events at center stage.

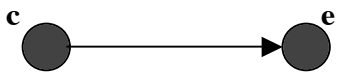

Figure 4.9

But it could have fired weakly—in which case the weak stimulatory signal sent to **e** would not have been enough to cause it to fire:

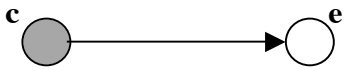

Figure 4.10

Some will want to say, about figure 4.9, that it is not **c**'s *firing* that causes **e** to fire, but rather **c**'s firing *intensely*. And some will take the further step of insisting that what we have in figure 4.9 are two distinct events, distinguished—despite their perfect coincidence in space, in time, and indeed in any categorical respect—by the fact that one could have occurred without the other, as witness figure 4.10. (Thus, the two events have different *essences*; see Yablo 1992.) I consider this strategy for accommodating the sorts of "fine-grained" causal distinctions exhibited by figures 4.9 and 4.10 both unattractive and unnecessary, and in chapter 7 will offer a very different approach. I thus explicitly reject the need to distinguish perfectly coincident events on the basis of their essences.

What I wish to point out here is merely that our account of causation can make no use of such distinctions, provided that the event "essences" are drawn solely from the *intrinsic* properties of the events in question. Suppose, that is, that two events *c* and *c** differ only in their essences, and this difference consists in the fact that it is essential to *c* that it have intrinsic property F, whereas it is only accidental to *c** that it have this property. (F might, for example, be the property of being an intense firing.) Thus, there is a possible world in which *c** but not *c* occurs (namely, a world in which *c** fails to have F), but no possible world in which *c* but not *c** occurs. Then consider a set S of co-occurring events that includes *c* but not *c**, and consider a corresponding set S* that is the result of replacing *c* in S with *c**. A minor problem immediately arises for our account of causation, since the crucial counterfactual that begins "if only the events in S had been occurring at the given time, then…" has an impossible antecedent: it cannot be that

all the events in S are occurring without it also being the case that all the events in S* are occurring, since *c* cannot occur without *c**. And a minor fix immediately suggests itself: understand the antecedent to be saying that no event wholly distinct from every event in S is occurring at the given time, where two events fail to be wholly distinct if, like *c* and *c**, they differ only in their essences.

Still, even with this obvious fix in hand our account will see no *causal* difference between *c* and *c**: for the counterfactual situation in which only the events in S are occurring will be *exactly the same* as the counterfactual situation in which only the events in S* are occurring, given that in constructing these situations we leave the events in the given set unchanged in intrinsic respects. Of course there are ways to introduce a causal difference, if we amend the account slightly. For example, we could return to the formulation that focused not merely on the counterfactual situation in which only the events in S occur at the given time, but on *all* nomologically possible situations in which this is the case. Since we would then travel beyond merely the "closest" worlds in which only the events in S occur, we would be sure to count as relevant to the causal status of *c** worlds in which it but not *c* is occurring. As noted, though, I will not explore this option, as I think there is a far better approach to handling fine-grained causation, and will develop it in chapter 7.

There is a second problem, which is that while we have finessed issues involving the exact occurrence conditions for the events in S, we have not yet avoided the need to provide occurrence conditions for *e*. Specifically, we need to know whether the counterfactual situation we construct in testing for the sufficiency of S is one in which *e* itself occurs, or a doppelganger of *e* occurs. Now, we can hope that in many cases, the counterfactual situation we construct will feature goings on in the time and place of *e*'s actual occurrence so similar to what actually happens there that there will be no doubt that *e* still occurs, in that situation. But not all cases will be like this. As an example, consider figure 4.11:

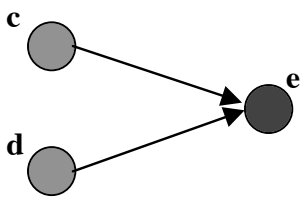

Figure 4.11

Let us stipulate here that **e** will fire if stimulated by **c** alone, but will not fire if stimulated by **d** alone. However, if, as here, both **c** and **d** fire, then **e** will fire very intensely, whereas it will only fire feebly if **c** alone stimulates it:

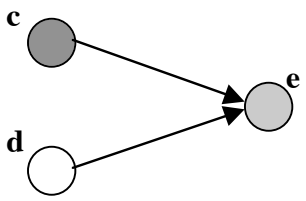

Figure 4.12

We might want to say, of figure 4.11, that the firing of **e** is jointly caused by *c* and *d*, the idea being that *c* alone is sufficient only for a feeble firing of **e**, and such a firing would be so different in manner from the actual firing of **e** as to constitute a numerically different event. On the other hand, we might want to say that *c* alone causes *e*, whereas *d* merely makes a difference to the manner of *e*'s occurrence. Which judgment we favor is certainly affected by how dramatic the difference is between the relevant situations. For example, if the occurrence of *d* only makes it the case that **e** fires at tiny bit more intensely, then we would be much more inclined to not count *d* a cause of *e*, but rather to relegate it to the status of something that merely makes a difference to how *e* occurs. But I also think that which judgment we favor can be heavily influenced by conversational context. In a way, that makes life easier, for it means that as far as understanding the metaphysics of causation is concerned, we do not need to provide the One True Account of the conditions of occurrence of *e*. We can it instead say that there are many acceptable versions of the occurrence conditions, with conversational context singling one (or a narrow range of them) out for attention. Again, though, I will postpone more detailed discussion of these issues until chapter 7.

§3 Two quick victories

We have arrived at the core of what I think is a successful regularity account of causation. Let us summarize. First, we have the foregoing definition of sufficiency. The definition of minimal sufficiency is then obvious: S is minimally sufficient for *e* just in case it is sufficient for *e*, but no proper subset of it is. And now we can produce our sufficient condition for causation: *c* is a cause of *e* if *c* precedes *e*, and, at the time of *c*'s occurrence, there is a unique minimally sufficient set for *e* that contains *c*.

The account needs a number of refinements, in particular so as to properly incorporate the tracing thesis. Before turning to them, I wish to record two victories.

§3.1 The problem of effects

First, we can now make the problem of effects disappear. Consider figure 4.2 again. Earlier, it looked as though *d* was bound to belong to a set of events minimally sufficient for *e*, and so, given that *b* likewise belongs to such a set—but a *different* set—our sufficient condition would fail to count *b* a cause of *e* (because the uniqueness condition was not met). But now we can see that this reasoning rested on a mistake. Let us first approach the issue in a way that respects the distinction between events and background conditions, folding a specification of these conditions into the definition of sufficiency. Then there are, at the relevant time t, only two events occurring: *d* and *b*. The set {*b*} is of course sufficient for *e*: for if *b* alone had occurred at t, then *e* (still) would have occurred (holding fixed the background conditions—in this case, the presence of the four neurons and the connections between them). But the set {*d*} is *not* sufficient for *e*, since had *d* alone occurred at t, *e* would *not* have occurred. So {*b*} is the sole minimally sufficient set for *e*, after all.

Do not complain that the imagined situations—the one in which **b** alone fires at t, and the one in which **d** alone fires at t—are nomologically impossible. In the first place, if that is a problem at all, it is just as much a problem for a counterfactual analysis (for such an analysis will need to consider what would have happened if **d** alone had not fired at t, and also what would have

happened if **b** alone had not fired at t). In the second place, it is not a problem at all, for there is no such nomological impossibility. Remember that the laws in question are the fundamental dynamical laws; not only are these compatible with the imagined counterfactual t-states, but it is perfectly clear and determinate how such states would evolve forward in time in accordance with such laws. (Cf. the discussion in chapter 2.)

Now let us run through the argument again, this time treating background circumstances as just consisting of so many events. Can we then get *d* to belong to a set minimally sufficient for *e*? No. For observe that our earlier choice—the set consisting of *d*, together with the events consisting in the presence of neurons **c**, **b**, and **e**, as well as the relevant stimulatory channels—is not sufficient for *e*. It does not follow from the fact that these events alone occur, together with the laws, that **e** fires. What is required in order to get the conclusion that **e** fires is that the set include the firing of **b**. But once it does so it is, while sufficient, not *minimally* sufficient: a proper subset obtained by removing *d* is likewise sufficient. So much for the first victory.

§3.2 Double prevention

For the second victory, let us recall from chapter 3 the kind of example that caused such headaches for counterfactual analyses: the distinctive sort of situation in which one event merely cancels a threat to another. Does our sufficient condition count such double prevention as a kind of causation? It does not. Consider for example figure 4.13:

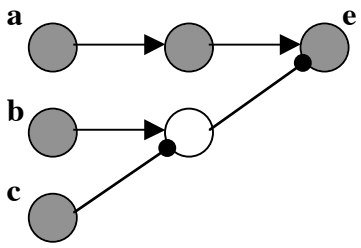

Figure 4.13

Here, the events occurring at the initial time t are just *a*, *b*, and *c* (once again treating the structure of the neuron network as part of the background circumstances, and not as being constituted by additional events; it is a simple exercise to confirm that this choice makes no

difference). We can now simply enumerate all the sets that can be formed from these events. Of these, the only set which is minimally sufficient for *e* is {*a*}. Note that this result is exactly what was wanted, given that our guiding intuition is that sufficiency is *aptness for producing*; while some sets that contain *c* are apt for producing *e*—namely, those that also contain *a*—*c* itself, intuitively, contributes *nothing* to this disposition. Remember, finally, that the headaches that beset counterfactual analyses when they were confronted with threat-cancelers were diagnostic of a point much deeper than that such analyses face possibly insuperable obstacles. There was, it emerged, no way to accommodate threat-cancelers as causes without having to give up too many important and basic claims about causation. The way to hold onto them, I argued, was to see them as characterizing a kind of causation distinct from—and perhaps more basic than—mere dependence. So this second victory is really quite a substantial one.

§4 The tracing thesis

So far, we have been focusing on one of the two theses that I said lay at the heart of the concept of production of that I am attempting to analyze. Time now to consider how the second thesis, what I have called the "tracing thesis", should be incorporated into this account.

§4.1 Transitivity, and easy cases of overdetermination

A simple and natural first thought is that we can incorporate the tracing thesis just by taking causation to be the ancestral of the condition we have already developed. Compare Lewis's early counterfactual analysis: immediately recognizing that counterfactual dependence between distinct events is not necessary for causation, he nevertheless took it to be sufficient for causation, and then analyzed causation more generally as the ancestral of this relation. Likewise, and for essentially the same reasons having to do with simple cases of overdetermination such as figure 4.4, we can recognize that it is not necessary for *c* to be a cause of *e* that it belong to a unique set of co-occurring events minimally sufficient for *e*. But if this condition succeeds as a sufficient condition, then perhaps we can take causation more generally simply to be its ancestral.

This move has an immediate payoff, for just as Lewis's account was, by this maneuver, able to handle simple cases of causal pre-emption, so too is our account. In figure 4.4, let time 0 be the time at which **a** and **c** fire, time 1 be the time at which **d** fires, and time 2 be the time at which **e** fires. True, at time 0 there are two distinct sets each minimally sufficient for *e*: {*c*} and {*a*}. But at time 1, there is just one such set: {*d*}. Moreover, there is at time 0 just one set minimally sufficient for *d*, namely {*c*}. Thus, our sufficient condition qualifies *c* as a cause of *d*, and *d* as a cause of *e*; taking the ancestral, we get the welcome result that *c* is a cause of *e*.

§4.2 Hard cases of overdetermination

Don't break out the champagne just yet. It should come as no surprise that, just as more stubborn cases of pre-emption foiled Lewis's original analysis, so too do they succeed in undermining ours. Consider the case where Suzy and Billy both throw rocks at a bottle, with perfect accuracy and deadly force, but Suzy's gets there first. Consider any time before the moment at which the bottle shatters. There will be, at that time, two distinct sets each minimally sufficient for the shattering: one containing either Suzy's throw or the appropriate event in the flight of her rock towards the bottle, the other containing either Billy's throw or the appropriate event in the flight of his rock towards the bottle. So it appears that our sufficient condition is never going to get any purchase. In the next chapter, will examine these and other hard cases of redundant causation in detail, partly with the aim of nailing down the case that the options for handling them are extremely limited. For now, I am simply going to jump ahead and sketch the one strategy that I think has any hope of handling these cases. Fortunately, it is also a strategy that handles them in an appealing and well-motivated manner.

First let us represent a typical case of late pre-emption more abstractly, as in figure 4.14:

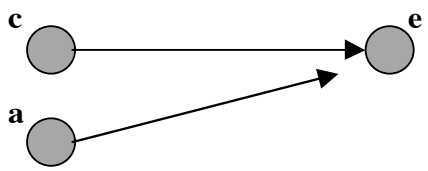

Figure 4.14

Here, neurons **a** and **c** fire simultaneously. But because neuron **a** is slightly farther away from neuron **e** than is **c**, its signal has not quite reached **e** by the time the **c**-signal stimulates **e** to fire (hence the arrow from **a** falls just short of **e**). As before, we will find, in the times before *e* occurs, no *unique* set of events minimally sufficient for *e*. But now consider that the events depicted in figure 4.14 contain within them an intrinsic duplicate of the events depicted in figure 4.15:

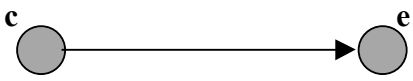

Figure 4.15

Suppose that you have come up with an account of causation that succeeds admirably in capturing the causal structure of figure 4.15. It then turns out to be possible to appeal to a thesis to the effect that the causal structure of a process is intrinsic to it in order to show that, since *c* in figure 4.15 is a cause of *e*, and since figure 4.14 contains within it an intrinsic duplicate of the events in figure 4.15, it must also be the case that in figure 4.14 *c* is a cause of *e*. I call this the "blueprint strategy" for handling overdetermination, for what we have in figure 4.15 is, as it were, a *blueprint* for part of what happens in figure 4.14. In general, the way to apply the strategy is as follows: Start with a stubborn case of overdetermination. Find within it an intrinsic duplicate of a much simpler case. Apply your analysis to this simpler case. If all goes well, that analysis will succeed in capturing the causal characteristics of the simpler case. Appeal to the intrinsicness thesis in order to "transfer" the causal characteristics of the simple blueprint over to the relevant parts of the more complex case with which you began.

It is a subtle and intricate matter how best to formulate the intrinsicness thesis that will effect such "transfers". Chapter 6 takes up this issue. Relevant here are two results of that discussion. The first result is that, unfortunately, the blueprint strategy does not extend to *every* case of redundant causation. It *does* extend to figure 4.4, for observe that we can "extract" from that case a much simpler one that we can expect our account to handle:

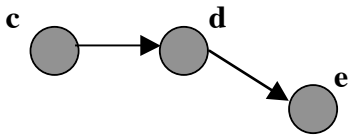

Figure 4.16

Since figure 4.4 contains an intrinsic duplicate of figure 4.16, the blueprint strategy will allow us to say that *c* in figure 4.4 is a cause of *e*. But there are more subtle ways that redundancy can come about. Consider for example figure 4.17:

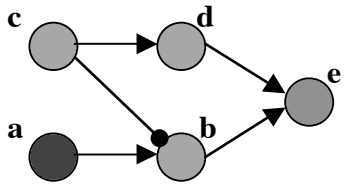

Figure 4.17

This case needs more explanation. This time, the inhibitory signal from **c** does not prevent **b** from firing, but merely reduces the strength of its firing—say, from 10 down to 5 (in whatever units). Simultaneously, the stimulatory signal from **c** causes **d** to fire with intensity 5. Finally, **e** is a stubborn neuron needing incoming stimulatory signals of combined strength 10 or greater in order to fire. Observe that at the time of **c**'s firing, there is a unique set minimally sufficient for *e*—but it does not contain *c*. It is just the set {*a*}. Our sufficient condition thus fails to count *c* a cause of *e*. Observe moreover that there is no way to handle this case by finding within it an intrinsic duplicate of a simpler case that our account can be expected to handle. (If this is not obvious, an explanation will be forthcoming in chapter 6.) So the blueprint strategy will not apply. On the other hand, our original move of "taking the ancestral" does help here, for {*c*} is the unique set minimally sufficient for *d*, and {*d*, *b*} is the unique set minimally sufficient for *e*. The upshot is that there are three kinds of redundant causation that we should distinguish: First, there are cases—such as cases of late pre-emption like figure 4.14—that only the blueprint strategy can handle. Second, there are cases—such as figure 4.4—that can be handled either by appeal to the blueprint strategy or by appeal to the transitivity of causation. Third, there are cases—such as figure 4.17—that can only be handled by appeal to the transitivity of causation.

For present purposes, the crucial point is that the development of the blueprint strategy will not allow us to dispense with an appeal to transitivity.

There is a second result the grounds for which will emerge more clearly in chapter 6. In order to make use of the blueprint strategy, an account of causation must be able to accurately map the causal characteristics of the simple cases I have called "blueprints". But it will turn out that it is not enough for the account to correctly identify all the causes of the target event *e* in such a blueprint. It must, in addition, be able to identify them as *being* all the causes. For example, in figure 4.15 it is not enough that the account correctly identify *c* and all the events making up the passage of the stimulatory signal to **e** as causes of *e*; it must also "say" that there are no *other* causes of *e* occurring within that time interval. But the problem is that no account that merely supplies a *sufficient* condition on causation will say any such thing.

You might think that we have here an intractable problem. For in order for an account to say, in addition, that there are no other causes of *e*, must it not also provide a *necessary* condition on causation? Well, no. For one thing, it might be that there *are no* other events that could serve as additional causes. For example, it might be that in figure 4.15 the only events occurring before *e* are the firing of **c** and the passage of the stimulatory signal to **e**. If so, then an analysis that correctly identifies these as causes of *e* can also safely add that there are no *other* causes of *e*, for the simple reason that there are no other events that *could* be causes. But it is too much to hope for that blueprints will always have this feature. For it may be unavoidable as a matter of fundamental law that the causes of *e* produce irrelevant side effects; for example, imagine that the stimulatory signal radiates energy as it travels towards **e**.

All the same, there is a way around our problem, if we can come up with an analysis that provides what I call a "turbo-sufficient" condition for causation. Such a condition would have the following abstract form: if circumstances are such-and-such, then the events in set S are all and only the causes of *e* occurring at the given time. In fact, our own sufficient condition on causation serves as an example of the form. That is, we could convert it to a turbo-sufficient condition by a slight amendment: say that when there is a unique set of co-occurring events

minimally sufficient for some later event *e*, then the events in that set are all and only the causes of *e* occurring at the given time. So we can see in the abstract that the problem set for us by the blueprint strategy—the problem of providing a test that will not merely identify all the causes of the target event *e*, but will identify them as *being* all the causes—is, in principle at least, soluble.

§4.3 Inadequacies of the sufficient condition

But we have not yet solved it, for two reasons: First, our sufficient condition on causation cannot be turned into a correct turbo-sufficient condition in the way just suggested. For consider figure 4.17. There, we see that at the initial time, there is a unique set minimally sufficient for *e*, namely {*a*}. But it fails to contain one of the causes of *e* occurring at that time, namely *c*. Therefore it is false to say in general that when there is a unique set of events minimally sufficient for some later event, then its members are all and only the causes of that later event occurring at the given time. That is one problem.

The second problem is that our sufficient condition is, in fact, incorrect, although it takes a rather arcane example to show why:

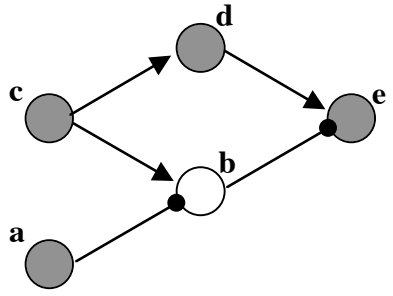

Figure 4.18

Here, **c** is a neuron that, if left to itself, will both attempt to stimulate **e** (via **d**) and will undermine this attempt (via **b**). As things happen, the firing of **a** blocks **c**'s attempt at self-undermining, and consequently **e** fires. My considered opinion about this example is that once we are careful to distinguish production from other kinds of causal relation, we should agreed that *c* is the sole producer of *e* occurring at the initial time; *a*, by contrast, merely prevents something that would have prevented *e*. But observe that at time 0 there is a unique set of events

minimally sufficient for *e* that includes *a*: it is the set {*c*, *a*}.[7] So our sufficient condition improperly judges *a* to be a producer of *e*.

§4.4 The sufficient condition refined and strengthened

We seem to have landed in a spot of trouble (you may be asking, "Who's 'we'?"). In the first place, it appears that we do not even have a successful sufficient condition on causation. In the second place, even if we did, we would seem unable to convert it into a successful turbo-sufficient condition. What to do? I think the key is to think more carefully about what the tracing thesis requires of an account of causation. Specifically, there are two ways that an account might fail to accord with that thesis. Suppose that event *e* occurs at time 2, and your account identifies some causes of it occurring at time 0, and some other causes of it occurring at time 1. Collect the time-0 causes into the set S_0, and the time-1 causes into the set S_1. Then one problem might be that S_0 fails to contain events that your account recognizes as causes of some of the events in S_1. That would be a failure of transitivity. But your account might also say, of some of the events in S_0, that they are not causes of any of the events in S_1. That would be a different problem: your account would be recognizing time-0 causes of *e* that, by its lights, do not bring about *e* by way of any time-1 causes of *e*. Either way, the tracing thesis—the idea that we find the causes of *e* by tracing back through successively less proximate causes of it—is violated. The thing to do, then, is to build it into the analysis that such violations cannot happen. In the case of our own analysis, here is the simplest and most direct way to do so:

Begin by supposing that the target effect *e* occurs at t', and that t is an earlier time such that *at each time* between t and t', there is a unique minimally sufficient set for *e*. Add the

[7] One could block this result by treating background conditions as events. For *a* only gets to belong to a set minimally sufficient for *e* because we hold fixed the presence of neuron **b** and the relevant connections between it, **c**, and **e**. If we treat the presence of neurons and the channels between them as events, then we can say that the sole set of time-0 events minimally sufficient for *e* contains *c*, the presence of neurons **d** and **e**, and the presence of the stimulatory channels connecting them. Still, I do not view this observation as providing any sort of refuge. It is, rather, an invitation to come up with examples with the same structure, but in which there are no such background conditions to appeal to. That is, we need only imagine a situation in which some event *c*, all by itself, initiates two processes, one of which will bring about *e* unless interrupted by the other.

requirement that whenever t_0 and t_1 are two such times ($t_0 < t_1$) and S_0 and S_1 the corresponding minimally sufficient sets, then

(i) for each member *c* of S_1, there is at t_0 a unique set minimally sufficient for *c*; and

(ii) the union of these minimally sufficient sets is S_0.

This requirement gives expression to the idea that when we, as it were, identify the producers of *e directly*, by appeal to their nomological relationship to *e*, we must get the same result as when we identify them by "tracing back" through intermediate producers.

We're now in a position to state the proper sufficient condition for production:

Given some event *e* occurring at time t' and given some earlier time t, say that *e* has a *pure causal history* back to time t just in case there is, at every time between t and t', a unique minimally sufficient set for *e*, and the collection of these sets meets the two foregoing constraints. Call the structure consisting of the members of these sets the "pure causal history" of *e*, back to time t; call the members themselves "proximate causes" of *e*. Then I claim that any proximate cause of *e* is a cause of *e*. Let us next take the ancestral of proximate causation, and call this relation "near-proximate causation"; then I claim that any near-proximate cause of *e* is a cause of *e*. And finally, I claim that the near-proximate causes of *e* back to some earlier time t are *all* the causes of *e* occurring between t and t'. Thus we have not only a sufficient condition on causation, but a turbo-sufficient condition.

With these revisions, our sufficient condition on causation can easily navigate examples such as figures 4.17 and 4.18. In figure 4.17, for instance, the events *b* and *d*, as well as all the events intermediate between them and *e*, will pass our test for proximate causation, whereas a more comprehensive stretch of *e*'s causal history that goes back to the time of *c* will not. No matter, since *a* and *c* clearly come out as near-proximate causes of *e*. In figure 4.18, *a* will neither come out as a proximate cause of *e*, nor will taking the ancestral connect it up to *e*. And finally, we can now safely take our condition on causation to be a turbo-sufficient condition. That is important, because while the analysis we have arrived that can handle a very wide range of cases, it cannot yet handle late pre-emption; we will need the blueprint strategy for that.

§5 Further refinements: the mereology of events

We must now address certain issues involving the metaphysics of events that arise when we recognize that events have both spatial and temporal mereological structure. Incorporating this fact into our analysis leads to some minor difficulties; let us consider them in turn.

§5.1 Events and their spatial parts

We have been assuming that events are momentary, lasting only for an instant, and for the time being let us stick with that assumption. All the same, they can be spread out in space, and thus be composed of distinct parts that are themselves events. Suppose then that one way of decomposing event c is into two non-overlapping parts c_1 and c_2. Let S be a set of co-occurring events to which c belongs. One immediate problem is that the antecedent of our counterfactual "had only the events in S been occurring at time t, …" won't be satisfied by any possible world. For a world in which the events in S are occurring is automatically a world in which the *parts* of c are occurring. But these parts are not identical to c itself. But this problem has an easy fix: understand the antecedent to require only that no event wholly distinct from every event in S is occurring, where two events count as wholly distinct just in case they share no part in common.

Suppose next that some set $\{c, \ldots\}$ containing c is minimally sufficient for some later event e. Then consider the set we get by replacing c with its two parts c_1 and c_2: $\{c_1, c_2, \ldots\}$. It might well be that this set is also minimally sufficient for e; that will be so if, intuitively, the occurrence of each of the two parts *matters* to whether e occurs. But even if not, at least one of the subsets we get by removing either c_1 or c_2 must be minimally sufficient for e. That follows from the fact that the set $\{c_1, c_2, \ldots\}$ is sufficient for e, but the set we get by removing *both* c_1 and c_2 is not. Either way, if c has more than one spatial part, then it *cannot* belong to a *unique* set minimally sufficient for e. At least one of the sets we get by replacing c with one or more of its parts must also be minimally sufficient for e.

Again, this is the sort of problem that strikes shallow, and asks only for a technical fix. To see what it is, let us indulge in a simplifying assumption, and take it that whenever there are

some momentary events *c, d,* …, all occurring at time t, there is in addition an event *c** which is their mereological fusion. Then we can say that such an event-fusion is itself sufficient for later event *e* just in case, were it alone occurring at the given time, then *e* would still occur—where it is understood that the counterfactual scenario is one in which all of the actual parts of the event-fusion are present. Put another way, we get to the counterfactual t-state by starting with the actual t-state, and stripping away everything that is not a part of *c**.

Intuitively, such an event-fusion might contain extraneous bits: there might be parts of it that aren't doing any work in making it the case that, in the counterfactual situation, *e* occurs. There is an obvious test for whether this is so: look to see whether some proper part of the event-fusion is itself sufficient for *e*. If this is *not* the case, then we will say that the event-fusion is minimally sufficient for *e*.

Now, remember why we introduced the uniqueness clause in the first place: we did that in order to ensure that our sufficient condition would not run afoul of overdetermination. Translated into the language of event-fusions, we can say that we have a case of overdetermination at time t just in case there are, at that time, two distinct event-fusions—distinct in the sense that one contains a part not had by the other—each minimally sufficient for *e*. Observe that when *e* is t-overdetermined, the two (or more) event-fusions that bear witness to this fact must *each* contain a part not had by the other; it is impossible that one be a proper part of the other (else the more comprehensive event-fusion would not be *minimally* sufficient). Suppose that *c** and *d** are two such event-fusions. Then there are, at time t, two distinct ways that *e* might come about—distinct because each essentially involves a part not had by the other.

Equipped with this notion of minimal sufficiency for event-fusions, we can now tidy up our account of minimal sufficiency for sets of garden variety events, and also tidy up the uniqueness clause. Let us say that a set of co-occurring events S is minimally sufficient for some later event *e* just in case the event-fusion formed out of the members of S is minimally sufficient for *e*. Let us say that S is uniquely minimally sufficient for *e* just in case there is, at the given time, no

other set S' that is minimally sufficient for *e*, and whose event-fusion is distinct from the event-fusion of the members of S.

That maneuver takes care of our first problem, although it leaves in its wake an annoying issue. Consider figure 4.15, a perfectly ordinary and simple case in which *c* causes *e*. We are imagining that *c*, the firing of neuron **c**, occurs all at once at time 0. But we are not imagining that it occurs at a spatial point: no, we can suppose that it takes place throughout a region the size of neuron **c**. It may well be that certain parts of this spatially extended event do no work in making it the case that *c* is capable of bringing about *e*. If so, then according to our new and stricter account, *c* itself will not qualify as a cause of *e*, but rather some *proper part* of it will. (Or it may turn out that *e* is symmetrically overdetermined, if there are different ways of carving up the spatial territory inhabited by *c* in order to get parts that are minimally sufficient for *e*. We will return to this point in the next chapter.) But in a simple case like this, it seems a mere annoyance to have to say that; one wants to reply that yes, *strictly speaking c* itself is not what causes *e*, but….

I think that as far as our ordinary judgments about what causes what are concerned, there are at least two different things going on here. The first is that we ordinarily ignore the spatial mereological structure of events. In analyzing what, at time t, causes some later event *e*, we carve up the goings on at t into wholly distinct events—wholly distinct, in the sense that they have no spatial parts in common. And then we happily proceed to ignore whatever spatial mereological structure these events have, pretending that they are, along that dimension, atomic, unless specific features of the situation force us to do otherwise.

That can easily happen, particularly if an event causes distinct outcomes by virtue of distinct parts that it possesses. For example, suppose that there is a battle, and soldier A and soldier B are both killed. As a first pass, we might happily say that a single event—the battle—caused two other events—the deaths. But then we look more closely, and see that soldier A died as a result of a skirmish that was happening over here, whereas soldier B died as a result of a skirmish that was happening over there. If so, then we would not want to say that the two deaths had a

common cause; at the very least, such a claim would misleadingly suggest that the case had a causal structure parallel to that depicted in figure 4.2. What would seem more accurate is to say that, strictly speaking, it was only one part of the battle that caused A's death, and a distinct part of the battle that caused B's death.

Still, it seems a safe assumption that in a very wide range of cases, no such considerations will force us to abandon the pretense that the events we are dealing with are, along spatial dimensions, mereologically atomic. Even when we do abandon this, a second sort of consideration can come into play that will allow us to say that, at least loosely speaking, a more comprehensive event causes some outcome, even though that event has irrelevant parts, and thus belongs to no set that passes our strict test for minimal sufficiency. Return to the case of the battle. Granted, if the question explicitly arises whether the two deaths had a common cause (at the time of the battle), we will say no, and by way of clarifying perhaps retract our initial claim that each was caused by the battle. But if we are speaking more loosely, we might be happy to credit the battle as a cause of each death *simply because* it had some part that (speaking more strictly now) caused the death. This kind of phenomenon, which I think is commonplace, might usefully be called a "touching the table" phenomenon, because of the parallels with the following sort of example: You put your hand on a table. Are you touching the table? Ordinarily we would say "yes" without hesitation. But then someone comes along and points out that *most* of the table's surface is *not* being touched by you. That *might* lead us to say that, "strictly speaking", you are only touching *part* of the table. But even if we choose to adopt such standards of strict speech, we will happily agree, at least in ordinary contexts, that it is enough to touch the table to touch one part of it.[8] Likewise, we will happily agree—again, at least in ordinary contexts—that it is enough for *c* to cause *e* for it to have one part that does so.

[8] Interestingly, other verbs seem to be trickier. For example, is it enough to *see* a table to see just one part of it—no matter how small, and regardless of whether the rest of the table is occluded? I don't know—though it seems fairly clear that such issues have little metaphysical import, being of more interest to those concerned with the semantics of such verbs.

Finally, it is worth noting that if we are happy to grant that an event *c* can cause a later event *e* merely in virtue of having a part that does so, then there is a simpler approach we can take to all of this. Suppose that we have two sets of events S and S*, each containing events that occur at some given time t. Suppose that S and S* are not the same. But suppose, finally, that we can find no event *c* in S and *c** in S* such that *c* and *c** are wholly distinct. Then we can simply stipulate that S and S* are not in competition, in the sense that our uniqueness clause, properly understood, is not violated, even if it is the case that S and S* meet our original condition for minimal sufficiency—that is, even if it true of both S and S*, but not of any proper subset of either, that had only the events in it been occurring at t, *e* still would have occurred. The idea is that if we have a case of *genuine* overdetermination of *e* by some events occurring at time t, then there must be some way of carving up the two "routes" to *e* into a set of t-events S and a distinct set of t-events S* such that at least one event in S is wholly distinct from at least one event in S*.

§5.2 Events and their temporal parts

Similar issues arise with respect to causal relations among temporally extended events. For purposes of understanding the "deep metaphysics" of causation, the best and cleanest assumption is that causation relates, in the first instance, only momentary events. What do I mean, "in the first instance"? Just that causal relations among all *other* events are going to be derivative on causal relations among momentary events (including, notably, the momentary temporal parts of temporally extended events). And why "cleanest"? Simply because the story about causal relations among momentary events can be laid out fairly precisely, whereas the story about how these causal relations fix causal relations more generally is bound to be messy.

Consider for example that there can easily be temporal analogs of our example of the battle and the two subsequent deaths. Imagine, for instance, a philosophy conference that takes place over a weekend. It has an early session, which Suzy attends, becoming annoyed as a result. And it has a much later session, which Billy attends, likewise becoming annoyed as a result. We could say that the conference annoyed Suzy, in virtue of having a temporal part that annoyed her;

and that it likewise and for the same reason annoyed Billy. But that would seem to be speaking loosely, as witness the impropriety of concluding that there was a *common cause* of Suzy's annoyance and Billy's annoyance. (It's worth remembering, in this regard, that the two sessions need not even be causally connected; it need not be the case that the early session is among the causes of the later session.) So we might want to say that strictly speaking, it wasn't really the *conference* that caused Suzy to become annoyed, but rather just one *part* of it, etc.

The same point applies on the end of effects: Suppose that the conference's early session is brought about as the result of the actions of one planning committee, whereas the later session is brought about as the result of a completely different planning committee. We might count the actions of each of these committees as among the causes of the conference itself. But it would be inaccurate to conclude that they *acted together* to help bring about the conference; at the very least, one would want to distinguish *this* kind of joint causation from the kind of joint causation exhibited in figure 4.17, where neuron **e** fires as the joint result of two incoming stimulatory signals.

Now, all of this is not to say that we can never apply our analysis of causation to temporally extended events. Of course we can; we must simply exercise a little bit of care. Suppose for example that *c* is some temporally extended event, and *e* is some later event—later in the sense that it begins to occur no sooner than *after* the onset of *c*. Then we might count *c* a cause of *e* if some temporal part of *c* qualifies, according to our analysis, as a cause of at least one temporal part of *e* occurring later than it—keeping in mind that, as the example of the conference shows, extenuating circumstances might lead us to retract this judgment.

§6 Causation involving omissions

§6.1 Causation by omission

It would seem that *omissions*—failures of events to occur, or perhaps more accurately, facts to the effect that no event of such-and-such type occurs—can play the role of causes. To be sure,

some philosophers deny appearances, often on the grounds that there are no such things as omissions to serve as causes. Helen Beebee provides an especially clear statement of the view:

> The reason I deny that there's any such thing as causation by absence is that I want to uphold the view that causation is a relation between events. To be rather more picturesque, I subscribe to what Helen Steward has recently dubbed the "network model of causation": the complete causal history of the universe can be represented by a sort of vast and mind-bogglingly complex "neuron diagram" of the kind commonly found in discussions of David Lewis, where the nodes represent events and the arrows between them represent causal relations. Or, to put it another way, the causal history of any event is, as Lewis puts it, "a relational structure".
>
> I think it's fair to say that the network model is the dominant model of causation in contemporary metaphysics—largely due, I guess, to the influence of Lewis and Davidson. The network model lies behind most of the theories of causation currently on offer; it also lies behind a vast amount of the literature on the philosophy of mind.
>
> If there *is* causation by absence, then the network model can't be right—or at least, it can't be right assuming there are no such things as negative events. And I assume in what follows that there *are* no negative events—which is to say, more or less, events whose essence is the absence of a property or particular. If Jones' failure to close the fire door is not an event, and if this failure was a cause of the fire, then the full causal history of the fire is not exhausted by the network of events and causal relations between them, for there will be no event of Jones' failure, and hence no causal relation between his failure and the fire. The network model cannot accommodate the fact—if it is a fact—that Jones' failure caused the fire, and hence cannot be the whole causal truth about reality.
>
> How should we solve the problem? One solution—the one I favour—is to hang on to the network model of causation and deny that there is any causation by absence .... (2003, p. XXX)

I do not wish to side with Beebee and her allies. The costs in terms of imputing such widespread error to our ordinary ways of thinking about causation seem too great. On the other hand, there are also philosophers who think there is rather a lot of causation by omission—who think, in particular, that all it takes for the failure of an event of some given type C to occur to cause some "positive" event *e* is for it to be the case that *had* an event of type C occurred, *e* would not have occurred. Here for example is Lewis:

> One reason for an aversion to causation by absences is that if there is any of it at all, there is a lot of it—far more of it than we would normally want to mention. At this very moment, we are being kept alive by an absence of nerve gas in the air we are breathing. The foe of causation by absences owes us an explanation of why we sometimes do say that an absence caused something. The friend of causation by absences owes us an explanation of why we sometimes refuse to say that an absence caused something, even when we have just the right pattern of dependence. I think the friend is much better able to pay his debt than the foe is to pay his. There are ever so many reasons why it might be inappropriate to say something true. It might be irrelevant to the conversation, it might convey a false hint, it might be known already to all concerned .... (2003, p. XXX)

This position also strikes me as absurd, for exactly the same reason: we quite systematically deny claims of causation by omission, even when they meet Lewis's counterfactual test. (As Sarah McGrath astutely pointed out to me, Lewis misleads here: for he says not that we *deny* such claims, but only that we "sometimes refuse to say" them. That omission on his part unfairly seduces the reader into thinking that his burden of explanation is much lighter than his opponents'.) Example: It may well be that the failure of my alarm to go off this morning caused me to be late in bringing my son to school; but I take it that it is something that we would intuitively count as *false*—and not merely something we would refuse to say—that Vladimir Putin's failure to wake me up on time likewise caused our tardiness, notwithstanding the fact that had he woken me up, we would not have been late. One might well wonder whether there is a safe middle ground between allowing too little and allowing too much causation by omission, and indeed, the question raises a host of fascinating issues that we will return to in chapter 9. For now, all I wish to point out is that there is no straightforward way to accommodate causation by omission within the framework developed in this chapter. That might seem a serious bug. In fact, I will argue in chapter 9 (with McGrath's help) that it is an attractive feature. Here I wish merely to emphasize that whatever its merits, it is also an *unavoidable* feature.

To see why, first observe that even if we treat omissions as a species of event—and even if we solve the considerable problem of how to locate omissions in time—it is almost trivial to show that no omission could ever belong to a set of co-occurring events minimally sufficient for some later event. For let omission *o* belong to set S, and let later event *e* be our candidate effect. Having understood omissions in general to be a kind of event, we face an immediate problem in trying to understand how there could be a world in which no event wholly distinct from those in S is occurring, at the given time t. And that is because to say that "nothing else is happening" would seem to imply that, in fact, *lots* else is happening—lots of *omissions*, that is. But let us fix this problem by taking it that when we require that no event wholly distinct from those in S be occurring, we mean no *positive* event—we don't mean to be quantifying over omissions, here. That fix will allow our definition of "sufficient" to apply coherently to S. But it is now

guaranteed that even if S is sufficient for *e*, it cannot be minimally sufficient. For the set S^- that results from removing *o* from S will likewise be sufficient for *e*, since a counterfactual circumstance in which no (positive) event wholly distinct from those in S is occurring will be exactly the same as a counterfactual circumstance in which no event wholly distinct from those in S^- is occurring. It follows that as far as our analysis is concerned, omissions can't even get into the running for being causes.

That is exactly as it should be. For remember that our analysis is intended to capture *production*—one central *kind* of causal relation, the kind of relation we have in mind when we ask what sequences of events, or what processes, *led up to* or *generated* some given effect. Omissions are not in the business of *generating* anything. Granted that they can, in some other and perfectly legitimate sense, *cause* outcomes, it is eminently reasonable to expect that this sense will require, by way of explication, a very different account from that appropriate for production.

Still, one might fairly wonder whether this result is an artifact of our account, whether some other account—perhaps a better one—still broadly in the spirit of regularity analyses could more readily accommodate causation by omission. I do not think this question can be settled decisively, but I think there are strong grounds for doubt, since the most obvious way of trying to amend our account in order to effect such accommodation fails. Suppose, that is, that instead of incorporating into our account of sufficiency a no-interference premise (by asking what would have happened if the events in some set had been the *only* events occurring at some given time), we simply include, as members of the set of events deemed sufficient for the later event *e*, various omissions. In figure 4.15, for example, we do not try to make it turn out that $\{c\}$ is sufficient for *e*. Rather, we let the set sufficient for *e* be $\{c, o_1, o_2, \ldots\}$, where each o_i is an *omission*—corresponding, presumably, to a distinct way that *c* might have been *prevented* from bringing about *e*. Put another way, it is not *c* alone that is sufficient for *e*, but rather *c* together with a vast collection of *absences*, where the "occurrence" of all those absences collectively amounts to the claim that nothing acted to prevent *c* from bringing about *e*.

Superficially promising, this maneuver nevertheless fails disastrously, even when we bracket various substantial worries about the ontological status of omissions. One problem, comparatively minor, is that there is no sensible way to place bounds on how many such omissions should be included—since there are, after all, countless ways that any given event might be prevented. That's unfortunate; it means, for example, that even in the most sanitized cases there will be no hope of providing anything like a formal test for sufficiency. (Compare the account I have offered: at least in the idealized cases represented by neuron diagrams, it is clear enough how one could formally demonstrate that a given set of events was sufficient (or not) for some other event.) But a much more serious problem is that once we proceed in this way, we will find, in *every* case, that our target effect *e* has multiple minimally sufficient sets. Hence, given the uniqueness requirement, no event will ever meet our sufficient condition for causation, and any hope of building a comprehensive analysis upon that condition will be dashed.

To see how this problem arises, consider the following diagram:

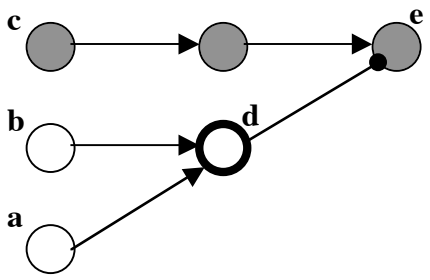

Figure 4.19

Here, **d** requires inputs from both **a** and **b** in order to fire. Since neither **a** nor **b** fires, the signal from **c** to **e** is, as it were, doubly safe. To what minimally sufficient set, then, does *c* belong? Ignoring the vast multitude of other omissions, there are at least these two: {*c*, *a**} and {*c*, *b**}, where *a** is **a**'s failure to fire (at the relevant time), and *b** **b**'s failure to fire. So *c* belongs to no *unique* minimally sufficient set for *e*. The problem generalizes: In any case where the action of some ordinary causes could have been disrupted by the *joint* occurrence of two events—one, say, of type A, and one of type B—and where no event of either type occurs, there will be at least two distinct sets minimally sufficient for the target effect, one containing as a member the omission

of an event of type A, and the other containing as a member the omission of an event of type B. And, of course, every set of ordinary causes is capable of disruption in some such way.

In short, then, it is not at all an accident that the sufficient condition for causation we arrived at cannot be satisfied by any omission. There was *no* good way to construct a regularity analysis that could straightforwardly accommodate this kind of causation. In addition, I think the reasons reveal a deep point about causation, and more specifically about the trouble one gets into if, in constructing a philosophical account of causation, one ignores the important differences between omissions and genuine events. Genuine events can stand in the interesting but rare causal relation of symmetric overdetermination, as in figure 4.20, where **c** and **d** both stimulate **e**, each with signals sufficient to cause **e** to fire all by itself:

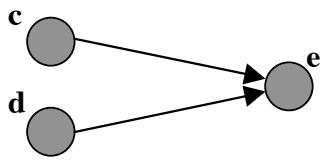

Figure 4.20

But if we really thought that omissions were just like ordinary, "positive" events in every causally relevant respect, then we should see symmetric overdetermination as absolutely commonplace, as figure 4.19 in effect shows.

Nor in fact would such commonplace symmetric overdetermination be confined to omissions. For *mixed* cases, in which a pair of an omission and an ordinary event "symmetrically overdetermine" some ordinary event would be just as commonplace. Consider that for any ordinary event *c*, and any later event *e*, one can always concoct some way that *c* could, if circumstances had been different, have acted to counteract some threat to *e*. As I am typing, war with Iraq is imminent. Let us suppose, depressingly, that it takes place a few months from now. Still, by typing I do not get any causal credit for the war, not even as a symmetrically overdetermining cause. But if omissions are to be treated just like ordinary events, and in particular are considered apt to join with ordinary events in standing in relations of symmetric overdetermination, then that intuitive judgment is false. For it takes just a little effort to imagine

various events occurring which would, in the imagined circumstances, have prevented war—but for my typing. (Let's cheat, and invoke a genie who is bent on stopping the war, unless she sees me typing; we can leave it as an exercise to get the same result, without cheating.) Why should we then not count the *omissions* of those would-be war-preventers, together with my typing, as collectively symmetrically overdetermining the war? Well, because to do so would be *mad*: by the same trick, we could show that *every* event is at least a symmetrically overdetermining cause of every later event. So the question should really be this: How are we to avoid this result?

The answer I favor, and which I will develop in the next two chapters, is that the causal phenomenon we intend to be marking off when we talk of "symmetric overdetermination" essentially involves *production*—and this is a kind of relation in which omissions cannot participate. It will follow that omissions *cannot* enter into relations of symmetric overdetermination, though we will see in chapter 9 that they can on rare occasion enter into relations easily mistaken for symmetric overdetermination. But all of these issues will become much clearer in the chapters ahead.

§6.2 Prevention

Let us turn finally to a brief consideration of *prevention*, which for the moment we will simply understand to be causation *of* omission. In chapter 9, I will argue that this is a mistake, and that a very different way of conceiving of prevention is called for. But for now, all I wish to point out is that, just as our account of production leaves no room for causation *by* omission, it also—modulo the provision of one crucial assumption—leaves no room for causation *of* omission. The crucial assumption concerns the fundamental laws, and it is this: A state of the world in which nothing whatsoever is happening will evolve, according to the laws, into future states of the world in which it remains the case that nothing is happening. For short, the empty state lawfully evolves into the empty state.

With this assumption in place, let us consider whether any set S of co-occurring events could ever be minimally sufficient for some later *omission o*. Apparently not. For by our assumption,

the *empty set* will be sufficient for *o*. If *o* simply consists in the failure of any event of some type C to occur at the later time t, then of course if nothing is happening at t, no event of type C is happening then, whence *o* is "happening" then. And if the empty set is sufficient for *o*, then of course no non-empty set can be *minimally* sufficient for *o*. The upshot is that there is no way, according to our analysis, for a set of events to "produce" some omission. Again, I think that is exactly as it should be. But that is *not* because I think that some other story should be told about causation of omission. In fact, I don't think that *any* story needs to be told about causation of omission, because I don't think it's an interesting topic. *Prevention*, to be sure, is an interesting topic—but as I will argue in chapter 9, the widespread equation in the literature of prevention with causation of omission is a mistake.

REFERENCES

Beebee, Helen 2003. “Causing and Nothingness”, in Collins, Hall, and Paul (eds) 2003.

Bennett, Jonathan 1988. Events and Their Names, Indianapolis: Hackett.

Collins, John, Ned Hall, and L. A. Paul (eds) 2003. *Causation and Counterfactuals,* Cambridge: M.I.T. Press.

Kim, Jaegwon 1971. “Causes and Events: Mackie on Causation”, Journal of Philosophy 68: 426-41.

Lewis, David 1973. “Causation”, Journal of Philosophy 70: 556-67. Reprinted in Lewis 1986: 159-172.

Lewis, David 1986. Philosophical Papers, Volume II, Oxford: Oxford University Press.

Lewis, David 1997. "Finkish Dispositions", Philosophical Quarterly 47: 143-158.

Lewis, David, 2003. “Void and Object”, in Collins, Hall, and Paul (eds) 2003.

Mellor, D.H. 2003. “For Facts as Causes and Effects”, in Collins, Hall, and Paul (eds) 2003.