2 | Causal and Explanatory Relevance

2.1 The Minimal Causal Account of Event Explanation

The theories of causation described in chapter one all admit a causal influence relation, and agree to a great extent on which events causally influence which other events. Causal influence is just the sort of relation to which the usual causal accounts of explanatory asymmetry appeal. Does an account of explanation—or at least, of deterministic event explanation, which will be my focus in this chapter—need anything more? Perhaps not. Perhaps an event is explained by whatever other events causally influence it, together with the laws and background conditions in virtue of which they do so. This is the *minimal causal account* of explanation.

The minimal account is a one-factor account that is not at all selective: to include every causal influence in the explanation of an event *e* is to include anything that, for example, exerts a gravitational influence on the objects involved in *e*. (To see this, consider the counterfactual test for causal influence: if the gravitational influences had not been present, then although *e* itself would still have occurred, it would have been realized by a slightly different concrete event, thus its actual concrete realizer would not have occurred.) Even the distant stars win a place in the explanation. Furthermore, the influence relation extends back in time without limit: the initial conditions of the big bang itself

are just as much causal influences on a present-day event *e* as the events in yesterday's news.[1]

The minimalist can censor much of this causal abundance by pointing to the negligible degree of influence exerted by most spatiotemporally distant events, a strategy discussed at the end of this section. But even then the minimalist account turns out to be insufficiently selective, as I will show in section 2.2. The remainder of the chapter explores various ways of supplementing the causal influence relation with a criterion for explanatory relevance so as to create a more selective explanatory relation, in the process moving from a one-factor to a two-factor account of causal explanation.

It is primarily as an expository device that I have introduced the minimal account is introduced, but I note that two well-known causal accounts of explanation are close in spirit to minimalism, those of Wesley Salmon and Peter Railton.

Salmon calls his explanatory causal relation *causal relevance*:

> . . . if we want to show why *e* occurred, we fill in the causally relevant processes and interactions that occupy the past light cone of *e* (Salmon 1984, 275).

Salmon's causal relevance is more or less what I am calling causal influence: one event is causally relevant to another

> . . . if there is a causal process connecting them, and if that causal process is responsible for the transmission of causal influence from one to the other (p. 207).

It seems, then, that Salmon counts all causal influences on an event, and nothing else, as explainers of an event, and so that he is a genuine minimalist.

1. For both Dowe and Lewis, causation is transitive by definition. For Woodward it is not, and transitivity fails for high level causation, but I conjecture Woodward causation between concrete events—in other words, Woodwardian causal influence—will always, or almost always, be transitive.

For Railton, not every explanation ought to be a causal explanation, but when a causal explanation is appropriate, it seems, no causal influence ought to be left out of the picture. Using the term *ideal text* to refer to the full explanation of an event, Railton writes that

> . . . an ideal text for the explanation of the outcome of a causal process would look something like this: an inter-connected series of law-based accounts of all the nodes and links in the causal network culminating in the explanandum, complete with a fully detailed description of the causal mechanisms involved and theoretical derivations of all the covering laws involved. This full-blown causal account would extend, via various relations of reduction and supervenience, to all levels of analysis, i.e., the ideal text would be closed under relations of causal dependence, reduction, and supervenience. It would be the whole story concerning why the explanandum occurred, relative to a correct theory of the lawful dependencies of the world. Such an ideal causal . . . text would be infinite if time were without beginning or infinitely divisible, and plainly there is no question of ever setting such an ideal text down on paper (Railton 1981, 247).

Whether Railton's notion of the explanatory causal relation is as liberal as my notion of causal influence is unclear, but Railtonian causal explanation certainly tends to inclusivity. Indeed, a Railtonian explanation will go beyond a minimal account in one respect: it will include information about "relations of reduction and supervenience" that I presume is not entailed by the facts about causal influence alone.

The most obvious difficulty facing the minimal causal account is the apparently unreasonable vastness of a complete minimal causal explanation. As I pointed out above, in a quasi-Newtonian world like our own, an event's minimal explanation ought in principle to mention anything that has ever exerted a gravitational force on the objects involved in the event, anything that

had previously exerted a force on these exerters, and so on. But all scientific explanations, even the most well regarded, describe much less than the complete causal history of the explanandum. The minimalist must make sense of this fact. Here it is possible to appeal to Railton's work in defending his own causal account.

First, consider the "ontological" sense of explanation, the sense in which an event's explanation is a set of scientific facts, as opposed to a communicative act (section 1.21). According to Railton, what is meant by the claim that science has discovered such an explanation is not that the ideal explanatory text has actually been created—that is more or less impossible—but that science is in a position to create the text. That is, all the knowledge and techniques required to create the text are in place; only the time, money and patience are lacking. Here is Railton again:

> The actual ideal is not to *produce* [ideal explanatory] texts, but to have the ability (in principle) to produce arbitrary parts of them. It is thus irrelevant whether individual scientists ever set out to fill in ideal texts as wholes, since within the division of labor among scientists it is possible to find someone (or, more precisely, some group) interested in developing the ability to fill in virtually any particular aspect of ideal texts—macro or micro, fundamental or "phenomenological", stretching over experimental or historical or geological or cosmological time (Railton 1981, 247).

What if science lacks the knowledge to construct a complete ideal explanatory text, but can construct some of the text? Railton's position is that science's explanation is good roughly in proportion to the amount of the ideal text that can be constructed (Railton 1981, 240–6). Every little detail makes the explanation a little better, but if a large piece of the ideal text cannot be constructed, science's explanation falls some way short of perfection.

Second, consider the sense in which an explanation is an act of communication. Railton holds that the quality of such an explanation is roughly

proportional to the amount of the ideal text that is transmitted to the audience. Since what is communicated is inevitably a vanishingly small fragment of the ideal text, an act of communication will always be, when considered on its intrinsic merits as a Railtonian explanation, a paltry thing, however useful it may be to its recipient. As the passage quoted above suggests, no one person ever comes close to understanding completely any phenomenon; if there is complete understanding, it is something possessed by the scientific community as a whole.

Pragmatics will, for Railton as much as anyone, play a role in the evaluation of explanations as acts of communication: an explainer ought not to be penalized for failing to mention an element of the ideal text that is common knowledge, for example, since you do not add to what is communicated by explicitly stating such things. But these pragmatic considerations are not proprietary to the study of explanation. Railton would, I think, like Lewis (1986a), deny that there is any distinctive pragmatics of explanation—correctly, I think.

In summary, then, according to the minimal causal account:

1. Every causal influence on an event is explanatorily relevant to its occurrence.

2. You do not fully understand a phenomenon until you are in a position (in principle) to articulate the role played in the production of the phenomenon by every one of these causal influences.

3. When you cannot construct the full history of causal influence, or you communicate less than the full history, the quality of the explanation increases with the proportion of the history you can construct, or that you do communicate.

As stated, the minimal account's requirements are demanding, but they can be used to defend the account against the charge that it mandates, absurdly, that the influence of the distant stars be included in the explanation of an everyday event such as a window's breaking. The defense has three steps.

First, a measure of *degree of causal influence* is introduced and the distant stars shown to have only a vanishingly small influence on the window's breaking. (I observed in section 1.43 that influence must in any case be quantified in order to deal with the barometer/storm asymmetry using facts about causal influence alone. See that discussion and note 11 of chapter 1 for two approaches to quantification.)

Second, the explanatory importance of a causal factor is equated with its degree of influence. It follows that the distant stars have almost no explanatory importance with respect to the window's breaking.

Finally, conversational pragmatics is invoked to argue that such unimportant factors will never appear in an explanation—in the communicative sense—of the window's breaking, as follows. Given the practical limits on the length of explanations, to mention the distant stars at all in an account of the breaking would be to implicitly ascribe to them a degree of importance far beyond their actual explanatory weight. Their relevance, though real, is simply so slight that in no imaginable context would it be conversationally correct even to allude to their presence. (Lewis (1986a) mentions some further pragmatic criteria that could plausibly be used to account for the stars' explanatory absence.)

None of this changes the fact that the stars are explanatorily relevant; what it does is to explain, rather well, the appearance of irrelevance. You may have lingering doubts: surely the stars are irrelevant not just in practice but in principle? I concur. But there are better places to take a stand.

2.2 The Problem of Relevance

2.21 Rasputin's Death

When a cabal of Russian nobles decided at last to do away with the mad Russian monk Rasputin, they took decisive action:

> . . . Rasputin was invited to visit Yusupov's home and, once there,

> was given poisoned wine and tea cakes. When he did not die, the frantic Yusupov shot him. Rasputin collapsed but was able to run out into the courtyard, where Purishkevich shot him again. The conspirators then bound him and threw him through a hole in the ice into the Neva River, where he finally died by drowning (Encyclopedia Britannica, CD 1998 standard edition, s.v. "Rasputin").

Rasputin's last stand makes a compelling story because, of all the death-promoting causal factors with which the conspirators assailed their victim—poison, shooting (twice), and drowning—the last alone explains Rasputin's death. No human can escape drowning when tossed through a hole in an icy river with hands and feet bound. Only a truly infallible method of murder could finish off Rasputin (though he is said by some to have partially freed himself before succumbing to the icy Neva).

On the minimalist account, all causal influences on an event, or at least all non-negligible causal influences, participate in that event's explanation. Rasputin's being poisoned and shot are causal influences on his death, influences that are, in contrast to the gravitational effect of distant stars, quite substantial. Consider: Rasputin is convulsing from the poison and bleeding from the gunshot wounds even as he finally dies of asphyxiation, so the process of his dying is thoroughly interwoven with the process of his convulsing and bleeding. Consequently, the poisoning and the shooting have a considerable effect on the concrete realization of his dying, and so count as major causal influences on the dying. The minimalist therefore has no choice but to accord them a large measure of explanatory importance.

According to the minimal account, in other words, poisoning and shooting are as much, or almost as much, a part of the explanation of Rasputin's death as his being bound up and thrown into the river. But this cannot be right. The poison and shooting failed to kill Rasputin, and so are irrelevant to his death, or at least, are vastly less relevant than his being thrown into the river.

Minimalism, it seems, is too minimal. What the enterprise of causal explanation requires is some criterion for distinguishing between the causal influences that explain Rasputin's death and those that do not. To find such a criterion is this chapter's *problem of explanatory relevance*. As you will recall from section 1.32, variants of the relevance problem have been noted many times. The seriousness of the problem, however, has often been underestimated, perhaps because the irrelevant factors appearing in the standard counterexamples have had so small a degree of causal influence—in many cases, such as the hexing of the salt, virtually none at all—that they are easily dealt with by the minimal account.

You may have noted a parallel between our explanatory and our causal commentary on Rasputin's death: just as we say that Rasputin's being thrown into the river, but not his being poisoned and shot, *explains* his death, so we say that his being thrown into the river, but not his being poisoned and shot, *caused* his death The causation in question is the kind of high level causal relation that the minimalist account deliberately ignores; the possibility of abandoning explanatory minimalism to appeal to high level causation will be discussed in due course (section 2.23).

2.22 Three Defenses of Minimalism

The relevance problem cannot be solved by even the most creative use of the resources allowed by the minimal account, or so I will argue, considering three minimalist strategies for dealing with the case of Rasputin. The first argues that, on the minimal account, Rasputin's influviation—his being bound and thrown into a river (cf. defenestration)—is, after all, a better explanation of his death than poisoning or shooting. The second tries to explain away the whole phenomenon of relevance by appealing to pragmatics. The third contends that poisoning and shooting are, contrary to the explanatory intuitions exploited above, highly explanatorily relevant to Rasputin's death.

To the first defense, then. The hope is to establish that an explanation that

cites Rasputin's influviation will contain more information about the causal network culminating in death than one that cites, say, poisoning. Clearly, the influviation and its consequences do not encompass a larger volume of the network than the poisoning; if anything, the reverse is true, since the poisoning occurs earlier and continues to exert its influence up until the moment of death. It would have to be, then, that the influviation has a greater degree of causal influence than the poisoning on the concrete realizer of the dying, and so that information about the influviation carries more explanatory weight.

It is plausible that this is so: more features of the dying's realizer depend on the influviation than depend on the poisoning. Thus the influviation is relatively more important, explanatorily speaking, than the poisoning. But this is not enough in itself to solve the relevance problem. What is shown is that the poisoning is somewhat less explanatory than the influviation, but what needs to be shown is that the poisoning is not relevant at all, or at least that it has minimal relevance.

Second, you might hope that the pragmatics of explanation will offer a way out. Could it be that poisoning is in fact just as relevant to Rasputin's death as influviation, objectively speaking—just as the minimalist account would have it—but that influviation appears to be far more important because it is more practically relevant in the conversational contexts where Rasputin's death is typically explained? If this were true, there would be a conversational context (perhaps very seldom realized) in which poisoning would be highly relevant, and therefore in which poisoning would have to be mentioned in an explanation. But there is no such context. Poisoning had little or nothing to do with Rasputin's death, so does not explain it; no context can alter this fact.

I do not deny that, in any particular act of explanation, context can make some factors more salient than others. But contextual salience—a pragmatic matter—is to be distinguished from explanatory relevance. Context can make salient or non-salient a factor that is already explanatorily relevant, but as the case of Rasputin's death shows, it cannot confer explanatory relevance where

it does not exist or remove it where it does exist. Poisoning is irrelevant to Rasputin's death and influviation is relevant, no matter what the context.

Of course, the conversational context can make poisoning relevant by altering the explanandum itself: if what is wanted is an explanation of why Rasputin's assassins resorted to throwing him into the river, their unsuccessful attempt at poisoning may well figure in the explanation. But that is simply to change the subject.

The third defense of minimalism dissents from the prevailing view that the explananda of event explanations are normally high level events (section 1.22), proposing instead that in scientific explanation, at least, they are usually or always concrete events, events individuated by every minute physical detail of their happening. The concrete event of Rasputin's death is an entity so fine-grained that it would have been a different event had the smallest detail of the death been different. The fact that Rasputin's body was full of poison or punctured by bullet holes is an essential, and not inconsiderable, part of the concrete event of Rasputin's death, hence it makes intuitive sense after all that poisoning and shooting should appear in the explanation of the fact that this event occurred—this *exact* event, in all its detail. In other words, if the explanation of Rasputin's death is an explanation of the death's concrete realizer, then poisoning and shooting are intuitively explanatorily relevant, just as the minimalist account implies.

As was pointed out long ago by Davidson (1967), following Hempel, this proposal does not sit well with our explanatory practice. Consider two propositions: that Rasputin died, and that Rasputin's body contained poison as he died. These two descriptions pick out the same concrete event, because the goings-on they name occupy the same spatio-temporal region. It follows that, if it is the concrete event that is the explanandum, then the two propositions, considered as descriptions of explananda, are equivalent, and so ought to attract the same explanation. But this is not so: the explanation of the first is that Rasputin was tied up and thrown into the river (poison is irrelevant), while the explanation

of the second is that Rasputin was poisoned shortly before his death (how he died, hence his being tied and thrown into the river, is irrelevant).

This last objection goes some way to exposing a fundamental flaw in the minimalist's way of thinking: according to minimalism, all explananda occurring or holding in the same pockets of space-time will have the same explanations. But as Rasputin's death and many other examples show, our explanatory practice is far more discriminating than the minimalist allows (see also Strevens (2003a, §4)). We distinguish between explananda whose occurrence depends on the content of the same space-time region in different ways, that is, between different high level events with the same concrete realizers, counting different causal influences as explanatorily relevant in each case. Treating explananda as high level events is an important prerequisite for all the accounts of explanatory relevance considered in this and the next chapter (section 3.3).

2.23 Augmenting the Minimal Account

The problem of explanatory relevance is the problem of picking out, from among all the causal influences on an event, those that genuinely explain the event. The task, then, might be neutrally characterized as follows: find a selection principle able to make the appropriate relevance distinctions, and then add it under some guise to the minimal causal account of explanation.

Although the development of an apt selection principle is still many pages away, it is not possible to avoid a major decision as to the strategy for incorporating the selection principle, whatever it may be, into the theory of explanation. Let me lay out the decision, and then show that it is not as final as it seems.

One way for an explanatory causalist to rein in the minimalist's judgments of explanatory relevance is to adopt a more selective conception of the causal relation. As I observed earlier, we are inclined to say that Rasputin's influviation, but not his being poisoned or shot, is a cause of his death. This would seem to indicate the existence of a high level causal relation between the influviation and the death, and the non-existence of a corresponding relation between the

poisoning or the shooting and the death.

High level causal relations of this sort are exactly what multilevel accounts of causation such as the counterfactual and manipulation accounts provide. In particular, the absence of a high level causal relation between, say, the poisoning and the death is indicated by the negative outcome of the counterfactual test for causal dependence between those two high level events: if Rasputin had not been poisoned, he would still have died.

The proper solution to the relevance problem for a causalist might seem clear, then. For the poisoning to explain Rasputin's death, it must not merely stand in the causal influence relation to the death; it must stand in a high level causal relation to the death. On the assumption that there is a high level causal relation between events *c* and *e* just in case the causal claim *c was a cause of e* is true, the solution gets you what you want: since we say that influviation was a cause of Rasputin's death whereas poisoning was not, the influviation but not the poisoning will qualify as a part of the explanation of the death. Whether Lewis's or Woodward's or some other view gives the correct truth conditions for causal claims does not matter; the idea is that whatever account of causal claims is correct, also solves the problem of explanatory relevance.

An alternative approach to the relevance problem—my own—aims not to invoke more causal metaphysics than is already inherent in the relation of causal influence. The selection rule that distinguishes the relevant causal influences from the rest is conceived not as a proper part of the metaphysics of causation, but as an independent element of a causal theory of explanation. I am proposing, then, what I called in section 1.1 a two-factor account of explanation. One factor is a causal relation, namely, causal influence. The other factor is a non-causal criterion of explanatory relevance that selects just those causal influences that are explanatorily relevant to a given explanandum.

The two-factor approach has the advantage of preserving metaphysical ecumenism: by committing itself to nothing more than the causal influence relation, it remains compatible with a broad range of views about the nature

of causation. This advantage, and the other advantages of two-factorism mentioned in section 1.1, may however seem to pale in the light of the following objection. We humans have a well-established practice of making causal claims such as *The poisoning did not cause Rasputin's death*. Unless we are deeply mistaken, this indicates the existence of, and our everyday deployment of, a wealth of high level causal relations eminently qualified to solve the relevance problem. Surely, even in the name of so high-minded a goal as ecumenism, it makes no sense to turn your back on such causal riches? Why not have it all: the high level *is a cause of* relation, a set of truth conditions for all of our causal claims, and a solution to the problem of explanatory relevance?

On my two-factor approach, I promise, you will have it all, in just as neat a package as the one-factorites proffer. Begin with the proposition that claims of the form *c was a cause of e* assert that *c* is a part of the causal explanation of *e*. This is, of course, a view necessarily endorsed by the kind of one-factor theory under consideration. But a two-factor theorist can—and I will—hold the same view.

The two-factor interpretation gives the thesis a new cast. The causal claim *c was a cause of e* does not, it turns out, assert the existence of a high level causal relation between *c* and *e*. Rather, it asserts the existence of an *explanatory relation* between the two events, the explanatory relation in question being a combination of a low level causal influence relation and the high level explanatory relevance relation.[2] Consequently, the correct account of explanatory relevance, when brought together with the relation of causal influence, gives you both the truth conditions for causal claims and an account of the high level *is a cause of* relation, now understood not as a purely causal relation but as the causal-explanatory relation, the relation that an event must bear to another event in order to participate as a cause in its explanation.

The fundamental difference between the two-factor and the one-factor views is, then, that on the two-factor view, the explanatory facts are prior to

2. A suggestion made some time ago by Davidson (1967, 160–1).

and account for our practice of making causal claims, whereas on the one-factor view, the relation expressed by causal claims is a high level causal relation that is prior to and forms the basis for our explanatory practice.

At this early stage, I do not expect you to assent to the two-factor over the one-factor approach. I do want you to see that the two-factor approach is not at an inherent disadvantage in giving a unified account of explanatory relevance, causal claims, and the appearance of a high level causal relation. These phenomena go together as neatly on a two-factor as on a one-factor approach.

How to decide, then, between the one-factor and the two-factor strategies? There seem to be limited grounds on which to opt for one over the other, yet the decision will determine, fundamentally, the character of causal explanation.

As it happens, although to make a decision at this point is convenient, it need not have momentous consequences. I will proceed as though the two-factor view is correct, searching for a criterion for explanatory relevance with which to sift through the information about causal influence. But in so doing, I will not entirely spurn the one-factorites. Any of the relevance criteria I consider in this study can be folded into the metaphysics of causation to yield a high level causal relation. That is, it is always possible to take a two-factor explanatory relation and to reinterpret it as a species of causal relation. This is true even for the explanatory relation I endorse, the kairetic difference-making relation. When *c* makes a difference to *e*, and so helps to explain *e*, say that there is a high level causal relation between *c* and *e*, if you will. Although, as the difference-making criterion soars to the heights of abstraction in parts three and four, it will seem less and less plausible that it plays any part in the metaphysics of causation, I do not forbid a metaphysical construal.

Further, both the one-factor and the two-factor strategies suggest the same difference-making tests for causal-explanatory relevance, namely, the probabilistic, counterfactual, manipulationist, and kairetic criteria examined in this and the next chapter. With a minimum of reinterpretation, then, my comments on

each are as pertinent to the one-factor as to the two-factor project. Though expository convenience calls for a schism, it need not go very deep.

2.3 The Probabilistic Solution

The criteria for explanatory relevance considered in what follows—the probabilistic, counterfactual, manipulation, and kairetic criteria—all pick out as explanatorily relevant those factors that, in some sense, *make a difference* to the fact that the explanandum obtains. But they offer quite different procedures for determining difference-makers.

The causal influences that make a difference to an explanandum, according to the probabilistic approach, are those that change the probability of the explanandum. (If the probability is decreased, an influence is negatively relevant. In what follows, I avoid the difficult question of the explanatory significance of negative relevance; however, I will take a stand in part four.)

Take the minimal causal account of explanation, or something like it, such as Salmon's account, and augment it with a probabilistic criterion for relevance. You then have the sort of probabilistic causal account of explanation advocated by Railton (1978), Humphreys (1981, 1989), and, in response to Hitchcock (1995)'s criticisms, Salmon (1997) himself. (Railton's probabilistic causal account is intended for a different class of explananda than the causal account sketched in section 2.1.) The equivalent one-factor approach posits a probabilistic account of high level causation, following such writers as Reichenbach (1956) and Suppes (1970).

Any probabilistic criterion for explanatory relevance that is applicable to Rasputin's death faces a striking prima facie problem: being poisoned generally increases your chance of dying. Thus Rasputin's being poisoned appears to pass the probabilistic test for explanatory relevance. (The same goes, of course, for his being shot.)

The probabilist's natural response to this problem is to hold that, although

being poisoned increases most people's chances of dying, it did not increase Rasputin's chance of dying. More exactly, being poisoned in this particular way and on this particular day did not affect his chance of dying, presumably because of some lucky combination of local factors—low level facts about the poison, the preparation of the tea cakes, Rasputin's metabolism, and his earlier meals that day—that rendered the poison ineffective.[3]

I will examine three objections to this strategy. The first objection is that, for a certain class of poisonings, the strategy either fails or trivializes the probabilistic relevance account.

The kind of cases I have in mind are those in which we would say that Rasputin's surviving the poisoning was a fluke. By this I mean that he survived not because of the presence of some systematically countervailing causal factor, such as an antidote or a particular kind of metabolism, but for one of the following two reasons:

1. Indeterministic case: some step of the poisoning process was indeterministic, and Rasputin was fortunate enough that the step happened to fall through.

2. Deterministic case: some step of the poisoning process depended on certain low level physiological details and, as it happened, that day the details were not quite right. In other words, the initial conditions deviated slightly from what was required for a successful poisoning.

I consider these scenarios in turn.

First, the indeterministic case. If the success of the poisoning turns on an indeterministic step, then there is no way for the probabilistic relevance account to avoid the judgment that poisoning was relevant to death. For in

3. An alternative response is to maintain that poison is irrelevant because it is screened off from Rasputin's death by his being thrown into the river (see section 1.47). This strategy will succeed only if influviation cannot fail to cause death—perhaps true in Rasputin's case, but obviously not true in any number of variant cases. An appeal to screening off will not do, then, as a general solution to the problem.

such a case, poisoning *did* raise the probability of Rasputin's dying, but he was lucky, and so he did not die. (For similar counterexamples to probabilistic approaches to relevance, see Achinstein (1983), §5.5 and Gluck and Gimbel (1997).)

The situation is analogous to the following unhistorical case, which has the advantage that the causal pathway is far simpler than in a poisoning. Suppose that Rasputin's enemies placed a bomb under his chair controlled by a quantum trigger that gave it a 75% probability of exploding. As it happened, the bomb did not explode, so they drowned him. Clearly, the placing of the bomb increased Rasputin's chances of dying; equally clearly, the bomb made no difference to his actual death. (This case is treated successfully by the kairetic account in section 11.3.)

Second, the deterministic case. In order for the probabilist to say that poisoning did not increase the probability of Rasputin's death in a scenario where a few apparently insignificant details made the difference between death and survival, every detail must be taken into account in fixing the probability for the death. But if every minute detail that might make a difference is taken into account when calculating probabilities, then given the assumption of underlying determinism, the probabilities can only come out as zero or one. Those factors that raise the probability of an event from zero to one will count as difference-makers; the rest will not. The result is something that resembles the counterfactual account of difference-making, to be discussed in the next section, more than any probabilistic relevance account.

In summary, the process by which poisoning leads to death is either indeterministic or not. If indeterministic, the probabilistic relevance account fails. If deterministic, in order not to fail, the account must take so much detail into account when determining probabilities that they all go to zero or one, in which case it is transformed into a kind of counterfactual account.

A second and related problem with the probabilistic handling of the Rasputin case: we do not know whether the poisoning process in Rasputin's case

was genuinely indeterministic or not, and if indeterministic, whether it was the kind of case where Rasputin's survival was due to brute good luck—where the poisoning raised the probability of death but he lived anyway—or where, because of the presence of some factor, such as a cautionary advance dose of the antidote, the poisoning never raised the probability of death at all. That is, we do not know whether or not the poisoning raised the probability of Rasputin's death. But we confidently judge that the poisoning was irrelevant to death. Therefore, the poisoning's irrelevance cannot turn on the question of whether death's probability was raised. In this case, at least, there is more to our judgment of relevance than a difference in probabilities.

Now to the third objection, the nub of which is not so much that the probabilistic relevance account fails to handle a particular kind of case, but that it is missing an essential component.

Suppose that the problem raised above, that of which details are to play a role in fixing the value of a probability, is solved, so that for any specification of a scenario and an event, there is a definite probability of the event's occurring in the given scenario. Then there is a fact of the matter about Rasputin's probability of dying after he is given the poison. This fact is not enough, however, to determine whether poisoning is probabilistically relevant to death. Also required is a fact about the probability of Rasputin's dying, had he not been given the poison.

In order to settle on such a probability, you need a procedure for determining the relevant counterfactual, non-poisoning scenario. In the Rasputin case, this is not so easy. Should your specification of the scenario mention the fact that Rasputin's assassins had him in their power and were determined to kill him? If so, it seems that no specific causal factor, not even influviation, will raise the probability of death, since that probability is already at its maximum value of one. The assassins' intentions, then, had better be left out. But what principle guides these decisions? The probabilistic relevance account of

difference-making does not provide an answer.[4]

Let me put this in the form of a general observation about any "difference-making" criterion for explanatory relevance. All such accounts have a common form. To determine whether a causal influence *c* makes a difference to an explanandum *e*, a comparison is made between two scenarios: the actual scenario, in which *c* is present, and a non-actual scenario in which *c* is not present. The facility with which *e* occurs in each scenario is evaluated. If it varies, then *c* is classified as a difference-maker. There are two steps, then, to the comparison:

1. The two scenarios to be compared are determined. The principal problem is to determine the details of the non-actual scenario, the scenario from which *c* has in some sense been removed.

2. The facility with which *e* occurs in each scenario is evaluated, and these "facilities" compared.

What counts as a scenario and what determines the "the facility with which *e* occurs" depends on the account of difference-making. In the probabilistic account, a scenario might be a model in some probabilistic theory, and the facility with which *e* occurs the probability that the model ascribes to *e*. In Lewis's counterfactual account, a scenario is a possible world, and the facility with which *e* occurs is the truth value, in the world, of the proposition that *e* occurred (no probabilities are involved in Lewis's basic account).

A complete difference-making account of explanatory relevance will, then, contain two parts: a removal procedure and a comparison procedure. The probabilistic criterion provides a comparison procedure—a procedure that will, once it is amended to deal with exploding chairs and the like, continue to be useful—but it lacks a removal procedure. The counterfactual and manipulation criteria, to be considered next, fill this gap.

4. Hitchcock (1993) discusses the parallel problem for probabilistic accounts of causal, as opposed to explanatory, relevance.

2.4 The Counterfactual Solution

According to what I will call the simple counterfactual approach to explanatory relevance, an event *c* that causally influences another event *e* is explanatorily relevant to *e* just in case, had *c* not occurred, *e* would not have occurred. Adopting such a criterion is, of course, roughly equivalent to taking the one-factor approach to causal explanation and adopting a simple counterfactual account of high level causation. (It is not quite equivalent, because there is no causal influence requirement for the one-factor approach's high level causation.) Lewis (1986a) himself exploits the similarity by advocating a one-factor account of causal explanation on which the explanatory high level causal relation is given, as you would expect, his own version of the counterfactual account.

The great virtue of a counterfactual approach to relevance is that, when conjoined with the Stalnaker/Lewis account of counterfactuals, it gives a full account of removal, thus satisfying the demand made at the end of the previous section. Let me elaborate this claim using the simple counterfactual criterion for relevance and Lewis's original account of counterfactuals' truth conditions (Lewis 1973b). The effect of removing a factor, on this approach, is to be determined by finding the closest possible worlds—the possible worlds most similar to the actual world in both matters of actual fact and in laws of nature—in which the factor is not present. Whatever holds in all such worlds is what holds if the factor is removed. (I ignore here the possibility, allowed by Lewis, that there is no maximally close world but a set of worlds of increasing closeness.)

The heart of Lewis's removal procedure is the account of the relation of closeness between possible worlds. When the factor to be removed is a spatiotemporally discrete event, such as a shooting or an influviation, there is a relatively straightforward algorithm for determining the closest worlds. They are the worlds that best conform to the following description: (a) they are identical to ours up until shortly before *c* actually occurred, (b) at which time a small divergence from actuality (perhaps a "small miracle") prevents, as con-

servatively as possible, the occurrence of *c*, after which (c) events unfold as prescribed by the laws of the actual world. (For a comprehensive discussion of the algorithm, and in particular of the nature of "conservative" divergences, see Bennett (2003).) As you will see shortly, the removal of other kinds of states of affairs may be a more involved.

Let me raise two problems for the counterfactual account of difference-making, considered as an account of explanatory relevance. The first is the well-known problem of preemption.

Writing about probabilistic relevance in the previous section, I suggested that, given Rasputin's assassins' determination to kill him and the fact that they had him entirely within their power, they were sure to kill him one way or another. But if this is true, then it seems that, if they had not thrown Rasputin into the river, he would have died anyway by some other nefarious means. Rasputin's being thrown into the river fails the counterfactual test for difference-making, then, and is therefore counted as explanatorily irrelevant. This conclusion is not in accord with our explanatory practice. The simple counterfactual account will, as explained in more detail in section 6.2, make the same mistake in any case of preemption, that is, in any case where there was a backup cause that would have brought about the explanandum if the actual cause had not.

Lewisians have tried to solve the preemption problem in a number of ways. Lewis's original reaction was to define causation as the ancestral of counterfactual dependence, so that *c* is a cause of *e* just in case there is a series of events $d_1, \ldots, d_n$, such that d_1 depends on *c*, each of the other *d*s depends on its predecessor, and *e* depends on d_n. It is now accepted that this solution does not correctly treat cases of what is somewhat cryptically called *late preemption*, of which the Rasputin case is an example. Try to find an effect of influviation—a d_i—on which Rasputin's death counterfactually depends; you cannot, because for any promising *d*, Rasputin's assassins will react to the non-occurrence of *d* by finding some other way to kill him.

This is, of course, only the beginning of the debate between Lewis and his critics, which has involved two amendments to the counterfactual account, Lewis (1986c) and Lewis (2000), the second a radical reformulation that attempts to analyze causal claims using nothing more than the relation of causal influence. I argue against the radical reformulation in Strevens (2003a). But for my purposes here, what is important is not that the counterfactual account fails outright, but that it stalls. This is enough reason to consider a new alternative.

The preemption problem is essentially a problem with the counterfactual approach's removal procedure: it fails to remove or otherwise neutralize backup causes, so rendering actual causes irrelevant. My second objection to the counterfactual approach also arises from a problem with removal: though Lewis's removal procedure generally yields a well-defined result when removing events, it is less satisfactory when removing ongoing states of affairs, background conditions, facts about objects' structure, and so on (cf. Field (2003), 448–450).

Suppose that you want to explain why a lump of sodium exploded when thrown into a pool of water. The correct explanation depends on sodium's having a loosely bound outer electron, which makes it susceptible to ionization. But in addition to sodium's loosely bound electron, many other facts about its atomic structure causally influence its behavior. Its number of neutrons, twelve in the only naturally occurring isotope, also counts as a causal influence on the explosion: the neutron number in part determines, for example, sodium's density, and so the rate with which it sinks. Neutron number does not, however, explain the explosion.

An account of explanatory relevance should make this distinction, counting the loosely bound electron as relevant, and neutron number irrelevant, to the explosion. The counterfactual criterion for relevance will do so by examining claims such as *If the sodium had not had twelve neutrons per atom, it would not have exploded*. On the Lewis approach, such a counterfactual is evaluated by finding a possible world (or set of worlds—this complication will not affect the argument) in which the exploding sodium has some number of neutrons

other than twelve. If the sodium explodes in this world, the neutron number is irrelevant; if not, it is relevant.

A world in which the sodium sample has a different number of neutrons, or in which its outer electron is not loosely bound, is a world with a radically different physics from ours (at least in the vicinity of the sample). To determine whether sodium explodes in such a place, you must fix all the important details of this physics and reason about its consequences. This is a far more difficult task, I will suggest, than discerning the explanatory irrelevance of neutron number and the relevance of loose binding, so it cannot be, as the counterfactual account proposes, the means by which we arrive at knowledge about explanatory relevance.

Start with neutron number. In the actual world, the twelve neutron isotope of sodium is, as noted above, the metal's only stable form. In the closest world where the neutron number is different, is the sodium stable? If not, its reaction with water will depend on its rate of decay and the chemistry of the decay products. If it is stable, what facts about the physics of atomic nuclei have been changed to ensure stability? Will these interfere with sodium's chemical properties, in particular its proclivity to react with water? Vexing questions.

Or take the relevance of the loosely bound outer electron. What is the physics of a sodium sample where the outer electron is bound more tightly? Are the rules of shell-filling changed, so that sodium's *p* shell fits seven electrons rather than six? It would then be an unreactive substance like neon. But perhaps this is the wrong change to make: it would require too fundamental a revision of quantum chemistry to allow an odd number of electrons to fill a shell. Make the *p* shell accommodate eight electrons, then. Now the outer electrons are all tightly bound, but sodium is a reactive substance like fluorine.

Perhaps you should forget about shell-filling and instead make the electromagnetic force stronger, leaving sodium's eleventh electron alone in the outer shell but tying it more tightly to the nucleus. It will now take more energy to dislodge the electron—but there is more available, because the attraction

between the hydrogen nuclei in the water and the electron is also stronger. Or could you tweak the physics of sodium while leaving unchanged the physics for the constituents of water? Who knows what would happen then?

Now, none of the foregoing considerations rule out the possibility that there is a determinate fact about the closest physics (or closest range of physics) in which sodium lacks various of its actual properties, nor that there is a determinate fact about whether or not sodium reacts explosively with water in worlds with such physics. Perhaps the counterfactual account does supply an answer to questions about the explanatory relevance of sodium's properties. But this could not possibly be the way that *we* answer such questions. I doubt that any expert on sodium's chemistry would venture a view on the issues raised above. By contrast, they are able to assert confidently that neutron number is irrelevant, and loose binding relevant, to sodium's reaction with water. Our scientific practice settles questions about relevance, then, without having to settle questions about the relative closeness of various physics to our own. The correct philosophical account of that practice will do likewise.

As I remarked above, the counterfactual account's difficulty with both questions discussed in this section arises, in different ways, from its removal procedure's requiring you to extrapolate entire possible worlds around the absences you posit. You may fiddle with the details of Rasputin's death, but you must evaluate the consequences of your fiddling in a world that contains the entire Russian court, malevolent and powerful conspirators included. Likewise, your tweaking of the sodium sample's properties must be made consistent with the complete underlying physics of the sample, with the dismaying result not that your question gets the wrong answer, but that it is rendered for practical purposes unanswerable.

The holism of Lewis's removal procedure is perhaps appropriate for elucidating the truth conditions of counterfactual claims, but it makes impossibly heavy going of picking out explanatorily relevant factors. The correct difference-making criterion for relevance will, I suggest, allow relevance to be

deduced from isolated models of relatively small parts of the workings of the world. The manipulation account, to be considered next, and the kairetic criterion, both satisfy this demand.

2.5 The Manipulationist Solution

Can another difference-making account with a counterfactual flavor, the manipulation view advocated by Woodward (2003) and others, improve on the simple counterfactual view and its Lewisian variants? I will answer this question with reference to the same two issues discussed in the previous section, preemption and the question of the explanatory relevance of structural properties such as neutron number.

As with the counterfactual account, I will be taking an account of high level causation and converting it into a criterion for explanatory relevance suitable for a two-factor approach to causal explanation. What happens, then, if you count as explanatorily relevant to an event *e* only those causal influences that the Woodward account counts as the "actual causes" of *e*? Let me answer this question by explaining Woodward's handling of cases of preemption.[5]

Consider a simplified Rasputin case involving some very hands-off conspirators. The mad monk is invited to tea, but the conspirators depart before he arrives, leaving poisoned teacakes on the table. Worrying that Rasputin may have recently eaten, they construct a backup trap as well. If the teacakes remain

5. For the complete account of event causation, see Woodward (2003, §2.7) and the work of Halpern, Pearl and Hitchcock, on which Woodward's approach is based (Pearl 2000; Halpern and Pearl 2005; Hitchcock 2001a,b).

Although in what follows I treat only Woodward's "actual causes" as explanatorily relevant, Woodward himself advocates a broader conception of the explanatory relation on which some events that were not actual causes of the explanandum nevertheless play a part in its explanation, on the grounds that they are potential causes—factors, knowledge of which would give you the power, in principle, to manipulate events like the explanandum. What implications this has for the explanatory relevance of, say, poisoning to the death of Rasputin, it is hard to say (Strevens in press a).

untouched for a certain period, the floor of the room opens up, dropping Rasputin into the Neva.

The Woodward treatment begins, always, with a causal graph showing the relevant type level causal relations between variables (figure 2.1A). All variables in the graph are binary, corresponding to a given event's occurrence or non-occurrence. When an event occurs, it triggers the event indicated by the "+" arrow; when it does not occur, it triggers the event indicated by the "–" arrow. (Where there are no arrows, there are no consequences.)

The causal graph is capable of being instantiated in different ways. Suppose that, in the actual scenario, Rasputin eats the teacakes and dies. Then the instantiated graph is as shown in figure 2.1B (events that are crossed out fail to occur, uncrossed events occur).

Suppose that Rasputin did not eat the teacakes? What would have happened then? The graph can be used to represent the evaluation of this counterfactual, as follows.[6] Flip the variable representing the eating of the teacakes so that the eating event no longer occurs, and then stand back and let the consequences propagate through the graph. The result is as shown in figure 2.1C: the non-occurrence of the eating causes the floor to open, so that Rasputin is influviated and subsequently dies. This corresponds to our intuitive evaluation of the counterfactual question: if Rasputin had not eaten the teacakes he would still have died. Because Rasputin dies either way, the simple counterfactual test for explanatory relevance does not count the teacake-eating that actually occurs as relevant to his death, a classic preemption counterexample to the simple counterfactual account of either relevance or causation. (This is the sort of case, incidentally, that Lewis's more sophisticated counterfactual criterion is easily able to handle.)

Woodward's account does better than the simple counterfactual test. To

6. Such graphs do not always provide the resources to evaluate counterfactuals in the Lewis manner, since they represent only one small part of any of the relevant closest possible worlds. In this case, however, the other regions of the relevant worlds may be safely ignored.

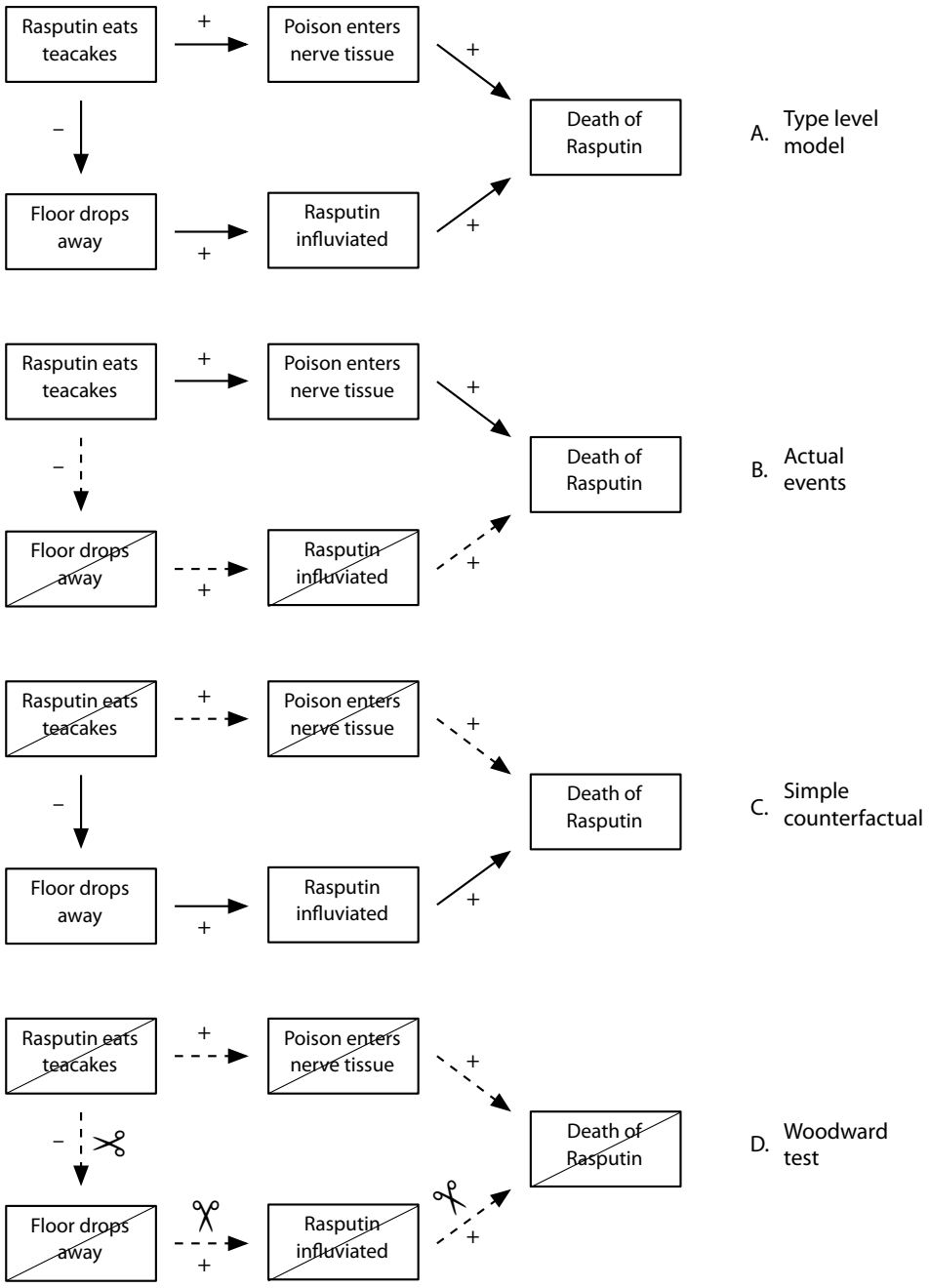

Figure 2.1: Counterfactual and manipulationist tests for difference-making contrasted

determine the relevance of an event *c* to another event *e*, a Woodwardian uses the following procedure:

1. Find a path in the (type level) causal graph leading from *c* to *e*, that is, a path that begins with a "+" arrow emerging from *c* and ends with a "+" arrow leading to *e*.

2. Assign all variables not on the path their actual values, then sever all the arrows pointing into these variables, in effect holding the variables constant no matter what (while maintaining any effects they may have on variables on the chosen path).

3. Flip the variable representing *c* so that it represents *c*'s non-occurrence, and let the effect propagate through the graph.

If this operation results in *e*'s not occurring for *any path* from *c* to *e*, then *c* is relevant to *e* (and causes *e*, on Woodward's interpretation). The relevance is, of course, by way of the path in question.

Figure 2.1D shows the Woodward manipulation of the graph for the only path leading from teacake-eating to death. The scissors indicate the notional cutting of the causal links not on this path. (You can see that in this case, just the first needs to be cut, but in principle they should all be cut.) When the eating is set not to occur, the floor remains in place, since the link between the non-eating and the floor's opening has been severed. Consequently, the death does not occur. For this reason, the death counterfactually depends in the Woodwardian sense on the eating, and so the eating is relevant to the death, as desired.

The Woodward account for the most part works well (for remaining problems, see sections 3.82 and 6.25), but it relies heavily on the discriminating power of the type level causal relations that connect the event variables. What are these relations? They cannot be relations of causal influence, for the following reason. The floor mechanism (so the story goes) is rather elaborate, and Rasputin studies it bemusedly as he dines on the teacakes. As a result, he eats

fewer teacakes than he might have otherwise, and dies a slower death. On this story, the floor mechanism is an important background condition influencing the way that the teacakes cause death—it makes a real difference to the way that death is realized. But then, if Woodward's arrows represent causal influence, the floor mechanism is on the putative causal path from teacake-eating to death, and so the link between teacakes and floor cannot be severed; as a consequence, when the model is manipulated so that the teacakes are no longer eaten, Rasputin will be deposited into the Neva.

It is crucial to the correct operation of the Woodward procedure, then, that irrelevant influences such as the floor mechanism are excluded at the type level. Intuitively, the causal generalizations represented in the graph with which the procedure begins must concern just those causal factors that can potentially make a difference to the event of interest, the explanandum.

What is the nature of these high level difference-making relations between variables? What is the metaphysical basis for the type level graph that tells all? Woodward sometimes appears to believe that the high level causal relations are metaphysically irreducible; at other times he appears to be agnostic on this question. But whether they are metaphysically reducible or not, there is something deeply unsatisfactory about the manipulationist's notion of relevance. On the one hand, if the high level relations between event types are irreducible—if they are, for example, metaphysically basic high level facts about manipulability—then the manipulationist relevance criterion is certainly well grounded, but at the cost of positing primitive facts about high level explanatory (or for a one-factor theorist, causal) relevance, a disappointing outcome. If on the other hand the high level relations are reducible—if they hold in virtue of lower level causal relations and other non-causal facts—the manipulationist account of explanatory (or causal) relevance is incomplete. In order to see what is and is not relevant to Rasputin's death, we need to know what does and does not lie on each of the various type level causal pathways to death. But Woodward and other manipulationists, whether out of a belief in irreducibility

or otherwise, stop short of supplying a reductive theory of type level causal relations. In short: the manipulationist may successfully account for high level relevance between singular events in terms of high level relevance between event types, but without any further story about relevance between types, the story is only half told.

Similar remarks apply to any manipulationist attempt at a solution to the problem, presented in the previous section, of sodium's reactivity. Somehow, it must be shown that reactivity does not depend on neutron number. It is plausible enough that no high level causal generalization exists relating reactivity to neutron number, whereas such a generalization does exist relating reactivity to the strength of sodium's outer electron's bond. Given these facts about what variables are related to what, the Woodward approach can account for the relevance of bond strength and the irrelevance of neutron number to a sodium-powered explosion. But as long as nothing is said as to why high level causal relations hold between some variables and not others, this hardly constitutes a solution to the relevance problem.

Part II

The Kairetic Account of Explanation

3 | The Kairetic Account of Difference-Making

3.1 Overview of the Kairetic Account

To understand a phenomenon is to see what made a difference to the causal production of the phenomenon and how it did so. At the heart of the kairetic account of explanation, then, is a criterion for difference-making. This chapter develops the criterion, and puts it to work determining explanatory relevance. Chapter four uses the relevance criterion to construct an account of event explanation. Chapter five extends the account of explanation in several important ways, and chapter six applies it to several outstanding problems concerning event explanation. Let me give you a preview of what is to come.

The kairetic theory provides a method for determining the aspects of a causal process that made a difference to the occurrence of a particular event. The essence of the theory is a procedure does the following: given as input a causal model M for the production of an event e, the procedure yields as output another causal model for e that contains only elements in M that made a difference to the production of e. A model that contains explanatory irrelevancies is, then, "distilled" so that it contains only explanatorily relevant factors.

An application of the kairetic procedure to M does not identify all the difference-makers for e, only difference-makers that appear in M. Further, the procedure may not identify all the difference-makers for e that are in M. I claim, however, that if a causal influence (a law, event, or background condition) is

a difference-maker for *e*, then there exists at least one causal model for *e* that represents the influence and with respect to which the kairetic procedure identifies the influence as a difference-maker. In principle, then, the method has the power to identify all the difference-makers for a given event.

How to show that a causal factor is not a difference-maker for *e*? Show—this is easier than it sounds—that the factor will be eliminated by the kairetic procedure from any causal model for *e* in which it appears.

The kairetic procedure functions as follows. I assume, in this initial treatment, that all causal models are deterministic. (The probabilistic case is taken up in part four.) A deterministic causal model for an event *e* is, I will suppose, a set of statements that, first, entails the occurrence of *e*, and second, does so in such a way (to be explained below) that the derivation of *e* mirrors the causal production of *e*. The difference-making parts of a causal model are those parts that play an essential role in the entailment, meaning roughly that, if they were to be removed from the model, it would no longer entail *e*'s occurrence, or would not do so in the right sort of way. This is what I call the *eliminative procedure*: remove as many pieces as you can from a model *M* for an event *e* without invalidating the entailment of *e*. Everything that remains in the model is a difference-maker for *e*. (Some precursors in the literature on explanation and high level causation, in particular Mill and Mackie, are discussed in section 3.83.)

The eliminative procedure turns out to be too crude a measure of difference-making in several ways. Once it has served its expository function, it will therefore be set aside in favor of what I call the *optimizing procedure* for determining difference-making. It is the optimizing procedure that will serve as the basis for the kairetic account of explanation developed in chapter four.

Whereas the first phase of the account of explanation, namely, the purging of explanatorily irrelevant factors from causal models, is essentially a matter of reduction, the second phase is a matter of construction. A *standalone explanation* of an event *e* is a causal model for *e* containing only difference-

makers for *e*. It is built from the models that have been stripped down by the eliminative or optimizing procedures. To understand the structure of standalone explanations is, I claim, the ultimate object of an inquiry into the nature of scientific explanation.

There are, I should note, many standalone explanations of any given event; these are all, in a sense, on a par, in that they are all complete and satisfactory scientific explanations of the event. But there are other senses in which one standalone explanation of an event may be better than another; these dimensions of explanatory goodness will be investigated in chapter four.

3.2 Causal Models

Explanatory information—that is, information about difference-making—is conveyed by a set of causal models that have been, first, stripped down by the kairetic procedure so as to contain only difference-makers, and then, sewn together to form a standalone explanation.

Before describing the kairetic procedure, I propose a canonical form for causal models, and I examine the way in which models of this form represent information both about the presence of causal influences—events, laws, and background conditions—and about the way in which these influences bear on a given event, hence about the causal production of that event.

A causal model for an event *e* is a representation of the different chains of causal influence that come together with the net effect of causally producing *e*. Such chains may be extremely complex—hence the metaphor of the causal web—and a detailed model will therefore be similarly complex, perhaps extending outwards to the distant stars and backwards to the beginning of time. Of practical necessity, any of our causal models will represent only a small part of a complete causal process. I will begin by considering the structure of the simplest class of causal models, those that represent a single link in a causal chain. I call these *atomic models*. More complex causal models are, as the name

suggests, constructed from atomic models.

3.21 Atomic Causal Models

A *veridical, deterministic, atomic causal model* for an event *e* is, for the purposes of this study, a set of true statements about the world that entail *e* (more precisely, the occurrence of *e*) in a certain way, to be specified shortly.

Such a model is veridical because the statements are true. It is deterministic because the statements entail that *e* occurs, rather than merely entailing that it occurs with some particular probability, or entailing a range of possibilities only one of which is *e*. (Non-deterministic causal models are discussed in section 9.7.) It is atomic because no intermediate steps in the causal process are identified explicitly. There may well be intermediate steps; what matters is that the steps are not spelled out in the model itself. Finally, the model is causal because the statements in the model do not merely entail that *e* occurs; they *causally entail e*, meaning that the entailment of *e*, or more exactly the derivation of *e*, mirrors a part of the causal process by which *e* was produced. The nature of this mirroring is the subject of section 3.23.

An atomic causal model for an event will have the same form as a DN explanation of that event:

> I threw a cannonball at the window,
>
> It is a law that the throwing of a cannonball at a window will cause the window to break, provided that nothing interferes with the ball's flight,
>
> Nothing interfered with the cannonball's flight, thus,
>
> ---
>
> The window was broken.

Both the causal model and the DN explanation are law-involving deductive arguments that the event occurred (using the term *law* liberally); the difference between them is that a causal model purports to represent a chain of causal influence running from the states of affairs identified by the premises to the event identified by the conclusion.

I call the event *e* whose occurrence is causally entailed by an atomic causal model the *target* of the model. I call the set of statements that make up the premises of the entailment the model's *setup*, and I call the derivation in virtue of which the entailment is demonstrated the model's *follow-through*. (A follow-through is, then, what Hempel in his presentation of the DN view called an argument, and what logicians call a proof.) It is the follow-through that must mirror the causal production of *e*, if the model is to *causally* entail *e*. The two kinds of information that must be conveyed by a causal model, then—the identity of the causal influences, and the process or processes by which their combined influence causally produced the target—reside in, respectively, the setup and the follow-through.

The simple atomic model for window-breaking contains, you will observe, a negative condition, requiring that nothing interfere with the flight of the cannonball. It is in general true that a causal model, when it cites enough detail to entail the occurrence of its target, will contain one or more negative conditions, in addition to laws, events, and other background conditions. Metaphysically, a negative condition can be understood, like any state of affairs, as a high level event (section 3.3). Thus it can be said without distortion that a causal model specifies only laws and events causally entailing its target. But negative conditions have a rather different feel than "positive" events such as cannonball hurlings and baseball bat swingings; though I insist on calling them "causal influences" (section 1.42), I acknowledge that the terminology is strained.

An atomic model may be thought of as representing a link in a causal chain, or perhaps better, a strand in a causal web, but it ought not to be thought of as the shortest possible link. Many events come between my throwing the cannonball and the window's breaking, and all of these events are a part of the causal chain leading to the breaking. The "link" corresponding to the above model, then, can be broken into many shorter links. Calling the model *atomic* is not intended to suggest otherwise. After all, in a continuous causal process

such as the breaking of the window, there is no shortest link. Conveniently, this frees up the term *atomic* for the use I put it to here.

A causal model is composed of statements, but in what language? Choose a natural language, and you limit the range of the model in whatever way that language is limited. It is crucial to the kairetic account of difference-making that the expressive power of the statements that make up a model be limited in no way whatsoever; I assume, then, an ideal language in which any state of affairs, law, or property can be represented. More or less equivalently, a model's setup might be taken to be composed of propositions rather than sentences.

3.22 Compound Causal Models

A compound causal model consists of two or more atomic models strung together to give a fuller description of the causal production of the target event. Take, for example, the model for window breaking above and add to it a model for my throwing the cannonball, say, the following derivation:

> I wanted to throw the cannonball at the window,
> Wanting to throw a thing invariably causes me to throw it, thus
> ---
> I threw the cannonball.

The result is a compound causal model for the window's breaking, in which the target of the new model, my throwing the cannonball, is a part of the setup for the original model, as shown in figure 3.1.

Like a veridical, deterministic, atomic causal model for *e*, a veridical, deterministic, compound causal model for *e* contains a set of true statements that causally entail *e*. In contrast to an atomic model, however, the compound model contains statements that are in effect intermediate steps in the derivation of *e*. My throwing the cannonball is, of course, the one such step in the window-breaking example. Intermediate steps in the derivation represent intermediate steps in the causal chain leading to *e*. I will call a statement corresponding to an intermediate step, or a set of such statements, an *intermediate setup*.

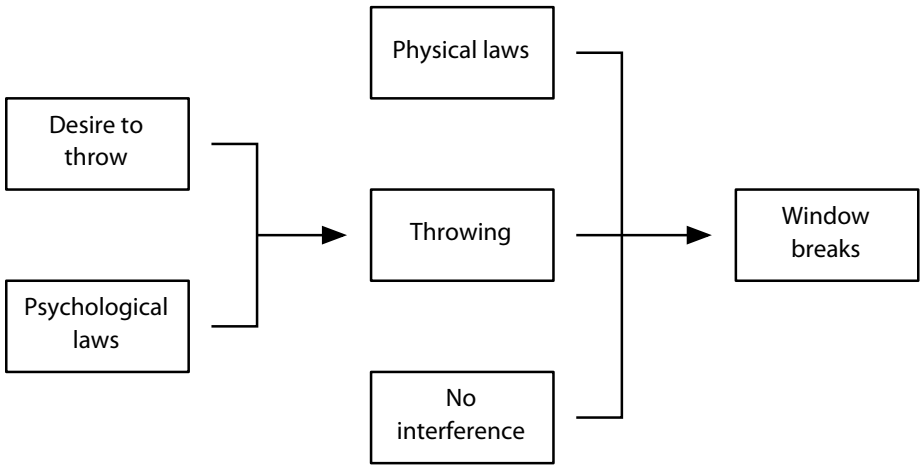

Figure 3.1: Compound causal model for a window's breaking

A compound model can be always be converted to an atomic model by removing its intermediate setups; the result is an atomic model that entails the compound model's intermediate setups. Thus a compound model does not convey any more information than its corresponding atomic model; it merely spells out some of what is already implicit in the atomic model. All the same, compound models have a distinct role to play in explanation; some of their uses are discussed in section 4.32.

3.23 Causal Entailment

The purpose of a causal model's follow-through, that is, the purpose of the logical derivation of the model's target from its setup, is to represent a causal process, namely, the process by which the various causal influences—events, laws, and background conditions—specified in the model's setup work together to causally produce the event that is the model's target. To perform this function, the follow-through must have the structure of an argument, or to put it more technically, a proof. In principle, then, every causal model consists of a setup and a proof of the target using the setup. In practice, the intended proof is usually obvious, and is omitted; in what follows I will often talk of

entailments without making explicit reference to the concomitant proofs.

When a model's follow-through is constructed so as to represent the causal production of its target, I say that its setup *causally entails* the target. An entailment's being causal is not a logical property, then, but rather a matter of its bearing a certain relation to the non-logical facts about causality, namely, its representing by the lights of those facts a possible causal pathway. How, then, does a derivation represent a causal process? What is causal entailment?

The occurrence of an event can be entailed by many different things. The breaking of a window, for example, might be entailed, as in the causal model above, by

1. The fact of my throwing a cannonball, some laws governing projectile motion and the molecular structure of glass, and the relevant background conditions.

But it might be equally well entailed by any of the following:

2. The content of a newspaper report describing the breaking and a generalization about the trustworthiness of the press,

3. A description of a photograph of my throwing the cannonball, a generalization concerning the accuracy of the camera, and the laws governing projectile motion and the molecular structure of glass together with the relevant background conditions, or

4. The conjunctive claim that the window broke and that Rasputin is dead.

You recognize immediately that the first of these entailments mirrors a causal process leading from the causal influences on the breaking to the breaking itself, whereas the others do not. Entailments (2) and (3) mirror causal processes involving the breaking, but they run counter to the direction of causation at some point. Entailment (4) does not represent a causal process at all.

The same recognition underlies the causalist's solution to the problems of explanatory asymmetry: in the DN explanations of the flagpole's height in

terms of the length of its shadow, and of the occurrence of the storm in terms of the barometer's dip, the direction of deduction runs either counter to the direction of causal influence or along a path where there is no non-negligible causal influence at all.

For my purpose—to characterize the notion of causal entailment—it is almost enough to rely on these intuitive judgments as to which entailments mirror causal processes and which do not; if you are prepared to accept the judgments at least provisionally, you might move on to the next section. If you have a skeptical turn of mind, however, you may suspect that our judgments about causal entailment are based in part on judgments about high level causal relations, or even worse, on judgments about causal-explanatory relevance. In what follows, I will show that an adequate notion of causal entailment can be built on the relation of causal influence alone.

Let me begin at the bottom, with the facts about causal entailment between concrete events. Consider a putative deterministic causal model for a concrete event *e* that cites in its setup another concrete event *c*. I assume that *c* is connected to *e* by a causal influence relation, with *c* as influencer. The question is whether the model's derivation of *e* mirrors the causal influence of *c* on *e*.

In the interests of clarity, I will make an assumption that is no doubt limiting in some respects: a causal influence relation between events *c* and *e* exists just in case there exists a causal law connecting a property *P* of *c* and a property *Q* of *e*, a law having roughly the form *If P*(*c*) *and Z then Q*(*e*), for some set of background conditions *Z*. (The formulation is clearly simplistic; a more realistic causal law will not refer directly to *c*, *e*, *P*, and *Q*; rather, it will be a universally quantified second order expression relating values of determinables, for example, relating the quantity of mass at one place to the component gravitational force due to that mass at another.)

What is a causal law? For the purposes of this discussion, it is a logical consequence of the fundamental laws that satisfies a further condition whose content is dictated by the correct metaphysics of causal influence. On the

conserved quantity view, a consequence of the fundamental laws of the form *If* $P(c)$ *and* Z *then* $Q(e)$ is a causal law just in case it holds because, in the presence of Z, there is a correctly oriented path of persistence and interaction running from the P-hood of c to the Q-hood of e. On the manipulation view, the condition is rather that the law holds because, in the presence of Z, there is a relation of manipulability between the relevant properties of c and e, in the sense that the P-hood of c can be used, in principle, to manipulate the Q-hood of e. Finally, on the counterfactual view, the causal law must hold because, in the presence of Z, e's Q-hood counterfactually depends on c's P-hood.

More generally, *If* $P(c)$ *and* Z *then* $Q(e)$ is a causal law if the fundamental laws together with the metaphysics of causal influence jointly imply that, if c has P and conditions Z hold, then e has Q in virtue of the causal influence of c's P-hood on e's Q-hood. (Again this is rough: because it is typically determinables that are related, such as component force and mass, it must be that the one determinable has its determinate value because of the causal influence of the other's determinate value.) Note that, first, this is not a definition but a definition schema whose content is determined by the correct metaphysics of causal influence, and second, that a causal law in my sense is not a fundamental but a derived entity. I should add that in section 7.6, I will propose a scheme for classifying causal generalizations that supersedes the notion of a causal law defined here (causal laws will turn out to lie partway between what I call "causal necessities" and "direct causal laws"); indeed, this notion will not figure anywhere but the definition of causal entailment.

I now define causal entailment in the obvious way: the derivation of e from c is a causal entailment just in case it goes by way of modus ponens applied to a causal law (or consists of a chain of such deductions).[1]

1. Strictly speaking, not every step in a causal entailment will represent a relation of causal influence. Using a typical system of natural deduction, for example, to get to e's Q-hood from c's P-hood by way of the causal law you will first have to apply a rule of and-introduction to get from $P(c)$ and Z to the antecedent of the law, the conjunction $P(c) \wedge Z$. Clearly the and-introduction does not represent a causal fact (or, I suppose, any non-trivial fact). Strictly

There is something more to say, as an example will show. Apply the characterization of causal entailment as it stands so far to a world of point mass particles in which the sole physics is Newtonian gravitation. Here, the gravitational force law is the sole causal law—or at least, it is a causal law once it is imbued with causal content by the correct metaphysics of causation, which will, I assume, identify each of two interacting particles as causally influencing the other in virtue of the gravitational force exerted.

The force law is not enough in itself, however, to determine how a particle's state (that is, its position and velocity) will change with time. You can use the force law to calculate the causal influence exerted on a particle by every other particle in the universe. But to infer the effects of the sum total of these influences you require a principle of *causal composition*, which states the net effect of the influences. The Newtonian principle of causal composition is of course very simple: the net force on, hence acceleration of, a particle is the vector sum of the individual component forces/accelerations.

All such composition principles have, I will assume, the same if/then form as the laws of influence, with the facts about individual influences as antecedents and the resulting net influence as consequent. A derivation that goes by way of a principle of causal composition is a causal entailment if it applies modus ponens to the principle in the obvious way.

I classify the composition principles, like the influence principles, as causal laws. All the causal laws in the Newtonian world, then, have (or can be put into) a natural if/then form, and causal entailment is a matter of deriving effects from causes by way of the straightforward application of modus ponens. I suggest that this holds true for almost any choice of causal metaphysics and fundamental physics. As the *almost any* indicates, what I have said above is not intended as an exhaustive theory of causal entailment, that is, as an exhaustive theory of the way in which science might use logical relations to represent

speaking, then, it is groups of steps, rather than individual steps, that must mirror a causal process in order for a derivation to constitute a causal entailment.

causal relations. But it is close enough to the way our current science represents causation with logic, I think, to provide the substrate for a theory of causal explanation.

This completes my criterion for determining when the derivation of one property instance from another is a causal entailment. What is wanted for a theory of causal models, however, is a criterion that applies to the derivation of events.

Begin with concrete events. That a concrete event occurs, or a concrete state of affairs obtains, is a matter of certain fundamental level properties being instantiated in a particular way, that is, in particular places and times. To derive the occurrence of a concrete event, then, is to derive the instantiation of some set of fundamental properties. Such a derivation is causal if the derivation of each of the property instances is causal. In other words, a derivation of the concrete event *e* mirrors the causation of *e* if it mirrors the causation of the property instantiations constitutive of *e*.

What, then, is a causal entailment of the occurrence of a high level event? For a set of concrete premises—premises concerning the occurrence of concrete events and the causal laws in virtue of which they have their influence—to causally entail the occurrence of a high level event *e* is for the premises to causally entail the occurrence of a concrete event that realizes *e*. For a set of high level premises—premises concerning the occurrence of high level events and so on—to causally entail the occurrence of a high level event *e* is for every concretization of the premises to causally entail *e* (where a concretization of a premise is a specification of some concrete realization of the premise).

Note that this talk of the concrete realizers of high level events, and of concretizations of high level premises, assumes the existence of correspondence rules linking higher and lower level vocabulary or properties. It is often said that correspondence rules will, or should, take the form of definitions, thus that they must give necessary and sufficient conditions for instantiating some high level property in fundamental level terms (Dupré 1993). You may doubt

that necessary and sufficient conditions of this sort exist. I doubt it too; it is important to see, therefore, that the kairetic account requires for its causal entailments something much weaker than this, namely, locally necessary conditions and locally sufficient conditions for high level property instantiation, of which there are more than enough. This liberal approach to correspondence will be defended and qualified in section 7.52; until then, I assume that the need for correspondence rules poses no special problems.

Let me show how the characterization of causal entailment applies to the window-breaking model above. For the model's derivation of the window's breaking to constitute a causal entailment, every concrete realization of the model must causally entail its target. The facts about realization are determined by the relevant correspondence rules, that is, the rules determining which concrete events constitute throwings, breakings, and so on.

Consider any one of the concrete models that realize the higher level model for window breaking. Such a model connects a concrete realization of the initial conditions specified in the setup with a concrete realization of the target. It represents, in other words, the way in which a maximally specific cannonball-throwing causes a maximally specific window-breaking.

The laws specified in the concrete model are, I will assume, the low level physical laws that connect the realizers of throwings, breakings, and so on; the sense in which these constitute a realization of the high level law cited in the high level model will be discussed shortly. According to any of the metaphysical accounts of causation introduced in chapter one, these laws are causal: they have an if/then form that relates an antecedent cause to a consequent effect, and they hold in virtue of the influence of the antecedent on the consequent.

Further, the concrete model's entailment of the breaking goes by way of the application of modus ponens to the low level if/then laws. For example, the deduction of the ball's trajectory from the fact of its being thrown will apply modus ponens to a causal law relating the force exerted on the ball by the thrower (on the "if" side) to its acceleration (on the "then" side). Or perhaps

it is better to understand the derivation as going by way of two applications of modus ponens to two laws: a law of influence relating the throwing to the component force, hence to the component acceleration, and a law of composition relating the component force and the absence of other significant forces to the net acceleration.

Three remarks on the example. First, in Newtonian and similar worlds, such as ours, an object's velocity at one time will count as a "causal influence" on its velocity at a later time. This is a straightforward consequence of the conserved quantity, manipulation, and counterfactual accounts of causation quite independently of the kairetic framework (because, for example, an object's position at one time typically depends counterfactually on its velocity at earlier times); it is, nevertheless, somewhat unintuitive. Perhaps a finer-grained account of our causal conceptual inventory would discriminate between "active" and "passive" effects. For the sake of simplicity, however, I am happy to go along with contemporary causal metaphysics' eliding of the distinction.

Second, the non-interference or no-other-forces aspect of the model's setup ("Nothing interfered with the cannonball's flight") is introduced by the principle of causal composition. I conjecture that in worlds with a broadly Newtonian structure to their mathematical physics (again I include our own), negative conditions are usually or always introduced in this way: they are assumptions of "no forces other than those specified" that are made in the application of the composition principle, in order to make the transition from an enumeration of certain component forces to the existence of a net force equal to the sum of the enumerated forces. Sometimes a negative condition may have the form "no other objects", but the forbidden objects will be those that, if they existed, would be in a position to exert forbidden forces. The topic of "non-interference" conditions is further discussed in section 6.22.

Third, as noted earlier, the causal laws that connect the throwing and the breaking in the concrete realization of the high level model must themselves be realizers, in some sense, of the the causal laws cited in the high level model,

which is to say, the high level law statements must be more abstract, less detailed characterizations of the same causal facts characterized by the low level law statements. In the window-breaking case, it is easy to see, because of the physicality of the high level laws, that this criterion is satisfied.

In biological or sociological explanation, the connection between high level and low level causal laws is far less clear. But the case that every high level causal law has a fundamental level realization will be made in section 7.6, where I argue that the nomological constituents of a high level law are certain perhaps abstract properties of the fundamental laws (see also section 3.52). This means that a high level law can be concretized as easily as a high level event: just find the fundamental laws, certain of whose properties constitute the high level law, or in other words, find the fundamental laws in virtue of which the high level law exists. These (along with entities I call basing generalizations, to be discussed in chapter seven) are the high level law's concrete realization. The notion of the concretization of a high level causal model does not founder, then, on the question of laws, and so my criterion for causal entailment in a high level model, which makes essential use of the notion of concretization, is sound.

The characterization of causal entailment offered in this section tightly binds the causality of a high level model to the causal facts at the fundamental level, as foreshadowed by my preference, expressed in chapter one, for invoking only fundamental level, as opposed to a multilevel, causal relations to resolve problems in the philosophy of explanation (section 1.4). In chapter one, I portrayed my view as ecumenical, on the grounds that many different views of the nature of causality (though not all) agree on the existence of fundamental level causal influence. To use fundamental level influence to resolve the explanatory asymmetries (the flagpole and the shadow; the gold foil and the scattering pattern) was therefore a conservative strategy. I am now proposing that the fundamental level provides sufficient metaphysical resources to un-

derstand any causal aspect of scientific explanation—that everything that can be explained causally, can be explained in terms of fundamental level causal influence, properly massaged.

This will strike many readers as strongly, perhaps radically, reductionist. It commits our explanatory practice to the thesis that all causality worth caring about for explanatory purposes has its sole origin in relations of fundamental level causal influence. Why think that the world will cooperate in providing what is needed? There are two connected worries to be addressed. (Note that in what follows, the reductionism with which I am concerned posits the reduction of high level causal facts to fundamental level causal facts; I will have nothing to say here or elsewhere on the question whether the fundamental level causal facts can be reduced to non-causal facts.)

First, some philosophers suspect that the thesis is in fact false: there are irreducible high level causal relations that are quite capable of providing understanding of the phenomena that they relate. Given what we now know, these suspicions are, I believe, extravagant: there simply are no causal relations of which we are aware that cannot be attributed to lower level interaction, and ultimately to the causal influence of fundamental particle on fundamental particle. This is a matter of empirical fact, a great lesson about the nature of the world learned from centuries of scientific work, not the conclusion of some a priori argument; it is not apodictic, then, but it can hardly be ignored. (I should add that the fundamental level causal story behind high level causal claims is not always straightforward, as is shown for example by the discussion of the causal role of omissions in section 6.3.)

Second, even if it is acknowledged that our world has a fundamental level that is the source of all (or at least, all scientifically significant) causality, it seems clear that things need not have turned out this way. Is it not implausible that our explanatory practice should have all along built such a strong assumption of causal reducibility into its being, an assumption that would make many apparently sensible possible worlds literally incomprehensible—their every

feature inexplicable—to us?

I might perhaps reply that many of our cognitive faculties are tailor-made for the actual world. But in truth, I believe that this objection to my explanatory reductionism is sound. In work published elsewhere, I have argued that our explanatory practice does not require the existence of a single fundamental causal level, but rather one or more domain-specific "basic levels" (Strevens 2007). How does this work? For any particular domain, we pre-theoretically suppose that there is a locally fundamental level, causality at which is the raw material of all explanation within the domain (and which constrains causal entailment just as the ultimate fundamental level does in the characterization above). Thus we may allow that there is a physically basic level, a biologically basic level, a psychologically basic level, and so on; the relations of causal influence that exist at these levels provide the metaphysical ingredients for explanations concerning the behavior of, respectively, physical objects, living things, and minds. The existence of multiple basic levels is quite compatible with anti-reductionism, and with the view that the world has no ultimately fundamental level at all. Thus our explanatory practice can make sense, in principle, of profoundly irreducible phenomena.

In our actual world, however, it has turned out (so we believe) that there is a single basic level, the fundamental physical level. Our practice has thus taken the form sketched above: all explanatory causality is derived from fundamental physics. For simplicity's sake, I am treating this as an axiom of our practice, when in fact it is derived from a more liberal metaphysical framework along with an observation about the causal structure of the actual world. In principle, our explanatory thinking is not reductionist, but in practice—in this world we inhabit—it is. My focus will be on the practice.

3.24 Other Approaches to Causal Modeling

Any system of representation that is able to furnish a setup and a follow-through—any system that can represent the fact that certain laws and states of

affairs obtained, and that can represent the structure of the causal dependence relations in virtue of which those laws and states of affairs causally produced an event—can be used, in principle, to construct a deterministic causal model with that event as its target. A diagram such as that shown in figure 3.1 above, for example, can be interpreted as a causal model in its own right, provided that the arrows are understood in the obvious way, that is, as relations of causal influence by which the states of affairs at one end jointly produce the event at the other end.

I use natural language sentences and entailment as my representational tools solely for convenience: natural language needs no further interpretation, and the entailment relation is familiar and well understood. But what is important, let me repeat, is not the means of representation but what is represented. What is represented are certain things in the world: states of affairs, laws, and the relations of causal influence between them. Any representation able to capture such facts can be used as a causal model, and indeed, there are a range of representational devices used in the scientific literature to communicate causal facts. By using a canonical form for my own causal models, I am in no way suggesting that this diversity of forms of causal representation is not to be desired.

3.3 States of Affairs

The explananda of event explanations are normally not concrete events, but entities that I call states of affairs or high level events, and that are sometimes called facts. States of affairs are important to explanation in three ways in addition to their role as explananda: they, rather than concrete events, are usually the difference-makers in an explanation, they have a role to play in regularity explanation that is just as important as their role in event explanation, and finally, they are akin to the difference-making relation itself in an important way.

What kinds of things might they be? Let me ask instead what work a metaphysics for states of affairs has to do, as far as the theory of explanation is concerned.

Two things. First, on a difference-making approach to the explanatory relevance criterion, what is important above all is that the states of affairs that appear in explanations have individuation conditions clear enough to determine whether they obtain or not. If you accept the simple counterfactual account of explanatory relevance, for example, to decide whether not an event *c* is relevant to an event *e*, you look to the closest world in which *c* does not occur to see whether or not *e* occurs. The outcome of such a procedure depends on the criteria that determine whether or not *c* and *e* occur in a given scenario or possible world. So there had better be clean individuation criteria for both the things that are to be explained and the things that do the explaining. Precisely how these criteria are to be determined, and what underlying metaphysics is to prop them up, is less important.

Second, the metaphysics of states of affairs ought to encompass any particular than might, in principle, be explained. This requirement inclines me towards the broadest possible answer to the question about the nature of states of affairs, an answer that has as a consequence that almost any event-like entity discussed in the philosophical literature is a species of state of affairs in my intended sense: concrete events, facts when distinguished from events, events in Kim (1973)'s sense, events in Quine (1960)'s sense, events in Davidson (1969)'s sense, negative states of affairs such as Zeus's not existing, and so on—since it would be bizarre to say that any of these entities is in principle inexplicable.

The desired range of states of affairs can be captured by a simple metaphysical schema. Let a state of affairs consist of two entities, standing in the instantiation relation. The first entity is some set of particulars: perhaps a region of space-time, or an object and an interval of time, or even the entire universe. The second entity is a property, perhaps relational. The state of affairs obtains, or equivalently, the high level event occurs, if the particulars instantiate

the property in the appropriate way. The event of Vesuvius erupting in 79 C.E., for example, might consist of a certain volcano and year (Vesuvius and the year 79), and the relational property of erupting in that year. The event occurred because the given particulars stand in the given relation: Vesuvius did erupt in 79.

Some remarks on this schema for states of affairs. First, a high level event is high level, rather than concrete, when the property whose instantiation is required for its occurrence is a relatively abstract property, a property that may be realized in a number of somewhat different ways. The event of Vesuvius's eruption is high level, for example, because the property of erupting in a given time span may be realized by eruptions that differ in many of their details. This event may be contrasted with the more concrete event of Vesuvius's having erupted in a very particular way, with the trajectory of the ash particles finely specified and so on, and at a very particular time. It is because of the high level event's relatively robust individuation conditions that many aspects of the geological process leading up to the eruption were not difference-makers for the eruption itself: they affected the way the eruption occurred, not the fact that there was an eruption.

Second, insofar as there are negative properties, there are negative states of affairs. These play an important part in explanation, most of all because various negative states of affairs must be satisfied (at least in our universe) for a causal process to run to completion. For a certain cannonball to break a certain window, for example, it may be necessary that nothing interfere (too much) with the cannonball's flight except the force of gravity. This state of affairs obtains just in case, for a certain period of time, an object—the cannonball—instantiates the property of being unaffected by any non-negligible non-gravitational forces. The negative state of affairs is composed, then, of the cannonball, the time frame, and the negative property just described.

Third, I use the terms *state of affairs* and *high level event* interchangeably,

choosing largely on the basis of whether it seems more appropriate to say that the thing *obtained* or *occurred*. I do not claim that these terms are synonymous, however; in fact, it seems likely that they are not. Arguably, for example, in everyday usage, the state of affairs of Rasputin's dying (in a certain time frame) and the event of his death—even conceived of as a high level, not a concrete event—are not the same thing. The state of affairs would have obtained had he died in absolutely any way you like; the event, however, perhaps has somewhat more fragile individuation conditions: had Rasputin been run down by a cart rather than murdered, the actual Rasputin death-event would not have occurred, though a different event, also a death of Rasputin, would have occurred in its place. (On these questions, and many more event-related subtleties, see Bennett (1988).)

This kind of issue I would like to avoid. It is easy to do so, because in almost any circumstances, the state of affairs and the high level event, even if distinct, will have the same explanation. Use the individuation criteria for the state of affairs, then, to determine the difference-makers for the death, and you will find the same difference-makers as had you used the individuation criteria for the event.

Fourth and finally, on the kairetic account, the difference-making relation itself turns out to fit the schema for a high level state of affairs. In the same way that breaking is a high level property of (some) windows, and a particular event of window-breaking is the instantiation of that property by a particular window, so the generic difference-making relation between a characteristic set of difference-makers for breaking and the generic breaking itself is a high level property of generic causal influence relations, and the obtaining of such a difference-making relation in a particular case is the instantiation of *that* property by a particular bundle of causal influences on a particular window-breaking.

3.4 The Eliminative Procedure

3.41 The Eliminative Procedure Characterized

The kairetic account understands explanatory relevance as causal difference-making, and it understands causal difference-making in deterministic systems as follows. A causal influence makes a difference to an event *e*—it plays an essential role in bringing about *e*—if there exists a veridical, deterministic, atomic causal model for *e* in which the influence plays an essential role in causally entailing *e*. Since the causal entailment represents the actual causal production of *e*, the thought goes, the factors essential for causal entailment are just those essential for causal production. Something like this proposal has been made in the context of theories of high level causation by Mackie (1974), following Mill (1973), and in the context of theories of high level explanation, though rather less explicitly, by Garfinkel (1981), following Putnam (1975). The primary challenge for difference-making accounts of this sort is to spell out what it is for a causal factor to play an essential role in a causal entailment. That task will occupy the remainder of the chapter.

Let me begin with a simple, rather obvious response to the challenge: a member of a set of propositions jointly entailing *e* is essential to the entailment just in case it cannot be removed from the set without the entailment's being destroyed. Never mind for now the complications that arise in certain cases, for example, if two propositions in the set are each sufficient in themselves to entail *e*; also ignore for now the sensitivity of such a criterion to the way in which information is packaged into propositions. These issues will be dealt with in due course.

When applied to the problem of determining difference-makers, the suggested criterion yields the following principle, a simple version of the kairetic criterion for difference-making:

> *If a factor c cannot be removed from a veridical, deterministic, atomic causal model for an event e without invalidating the entailment of e, then*

c is a difference-maker for e.

The converse does not hold: if *c* can be removed from a model for *e* without invalidating *e*'s entailment, *c* might nevertheless be a difference-maker for *e*. In order to establish that *c* is not among *e*'s difference-makers, it is necessary to show that *c* can be safely removed from *every* veridical, deterministic model for *e* in which it appears.

What does it mean to remove a causal factor *c* from a model? The factor must, of course, be a part of the model's setup, in the sense that one of the sentences in the setup must assert that *c* is present. To remove *c* from the model is to remove this sentence from the setup. The result is a model that is silent as to whether or not *c* was present—not, note, a model that asserts that *c* was absent. Removing *c* from a model is in this respect quite unlike moving to a possible world in which *c* is absent (Lewis's removal), or performing a Woodward manipulation as a result of which *c* no longer obtains (Woodward's removal). For Lewis and Woodward a scenario with *c* removed is one in which *c* determinately fails to occur; on the kairetic account it is rather one in which there is no fact of the matter as to whether *c* occurs.

It will be convenient to operate with a criterion for difference-making that, rather than defining what it is for a single factor to be a difference-maker, as does the principle articulated above, instead takes the form of a procedure to be applied to a causal model, the result of which is to pare down the model so that it contains only factors that are difference-makers for the target. The procedure, more exactly, takes a veridical, deterministic, atomic causal model for an event *e*, and produces another veridical, deterministic causal model for *e* with a setup that is a subset of the original model's setup, and in which all factors are difference-makers for *e*. I call this new atomic model an *explanatory kernel* for *e*. A causal factor is a difference-maker for *e*, then, just in case it appears in an explanatory kernel for *e*.

As the name suggests, an explanatory kernel for *e* is an explanation of *e*; furthermore, all atomic causal models that are explanations of *e* are explanatory

kernels for *e*. I will propose in section 4.1 that all compound causal models that are explanations of *e* are composed of explanatory kernels, either for *e* itself or for *e*'s difference-makers.

In what follows, I will describe several increasingly sophisticated procedures for extracting explanatory kernels from models. The first of these kernel determination procedures, based straightforwardly on the principle above, I call the *eliminative procedure*. It is quite simple: remove from a given model every causal factor that you can without invalidating the setup's entailment of the target, then stop. Everything that is left in the model is a difference-maker for the target. Let me now show how the eliminative procedure is able to deal with the issues that caused problems for other difference-making criteria in chapter two.

3.42 The Influence of Mars

Let me warm up with a ridiculously simple case, one that causes few problems for any account of difference-making. Throughout the course of events leading to Rasputin's death, the planet Mars exerted a slight gravitational pull on the principal actors. This pull is a bone fide causal influence on the death, according to any of the theories of causal influence described in chapter one, since Mars did have some effect, however small, on the concrete realization of the death. There will therefore be veridical atomic causal models for the death that specify Mars' gravitational effect. What I need to show is that such specifications are invariably removed by the eliminative procedure.

It is easy to see that this is the case. The elements in a causal model that do the work of entailing the death all concern ropes, ice, and rivers. Discourse on Mars can therefore be dropped without invalidating the entailment. What will be left behind is a preexisting negative condition specifying that there were no significant gravitational influences on the scenario except for Earth's.[2] Mars'

2. I am assuming that the negative condition is a part of any model that entails death, regardless of whether it mentions Mars. This might not be true of a model (which could never

influence, then, made no difference to the death.

Now some fine print. It is quite permissible to say informally that Mars made no difference to the death; strictly speaking, however, it is not Mars that is removed from the model by the eliminative procedure but the proposition that Mars has a certain property. What fails to make a difference to Rasputin's death, then, is not Mars itself, but the state of affairs of Mars' exerting such-and-such a causal influence. More generally, as foreshadowed above, difference-makers and non-difference-makers alike are, on the kairetic account, events or states of affairs—not objects or systems, but the fact of objects or systems instantiating certain properties.

Further, it would be an error to think that the application of the eliminative procedure produces a model that has nothing whatsoever to say about Mars. The model, by specifying an upper limit on the total non-terrestrial gravitational influence, places an implicit limit on Mars' gravitational field in particular, and thus implicitly attributes a property to Mars. This property, since it remains in the model after elimination, is a difference-maker: it makes a difference to Rasputin's death that Mars' gravitational pull was not enormously large, or to state the difference-making state of affairs more positively, that the pull fell into an interval with a certain upper bound. For more on this kind of negative difference-making, see section 6.3.

3.43 The Influence of Poisoning

Next consider Rasputin's poisoning. Although poison is a typical cause of death, in this particular case it does not explain the death; I need to show that the kairetic account concurs.

Take a veridical deterministic causal model for Rasputin's death that represents his being poisoned. Provided that the model also represents his being

exist in practice) that specifies every gravitational influence in detail. In such a model, the removal of Mars will be a part of an abstraction operation that produces such a condition; see section 3.52.

thrown into the river, the poisoning can be removed without invalidating the entailment of death. Poisoning, therefore, is not a difference-maker in virtue of such a model.

But what if Rasputin's influviation is not in the model? Can you construct a veridical model that causally entails Rasputin's death in a way that essentially involves his poisoning? If so, then because poisoning could not be removed from such a model, it would count as a difference-maker. It had better be that any model for Rasputin's death that hinges on his poisoning is either not veridical or does not genuinely entail his death.

What would be the structure of a model for Rasputin's death by poisoning? Very schematically, its setup would have two parts: initial conditions specifying that Rasputin was poisoned in such and such a way and that other important background conditions *Z* held, and laws having a consequence of the following form:

> When a person is poisoned in such and such a way, and conditions *Z* obtain, that person will die.

(This generalization connecting poisoning and death need not itself be a law; see section 3.46.)

In order to entail death, the poisoning model must specify that conditions *Z* held. But conditions *Z* could not have held, because Rasputin *was* poisoned in such and such a way, yet he did not succumb. Whatever conditions were necessary for a successful poisoning were not wholly present. If the poisoning model asserts that these success conditions held, it is not veridical; if it fails to assert that the conditions held, it does not causally entail death. Thus there is no model in virtue of which poison counts as a difference-maker for Rasputin's death. (For the kairetic account's treatment of a modified Rasputin story in which the poison would have eventually killed him, had he not been drowned first, see the treatment of "late preemption" in section 6.23.)

3.44 The Influence of Influviation

Finally, consider Rasputin's influviation, that is, his being bound and thrown into a river. Influviation made a difference to Rasputin's death; I need to show that the kairetic criterion counts it as a difference-maker. To this end, consider a deterministic causal model for Rasputin's death by influviation constructed along the same lines as the model for death by poisoning described above. The model's setup states that Rasputin was bound and thrown into a river, that various other background conditions obtained, and that it is a consequence of various relevant laws that people bound and thrown into a river in the stated conditions die. If constructed correctly, the model is veridical and entails Rasputin's death; further, influviation cannot be removed from the model without invalidating the entailment. Influviation, then, is a difference-maker.

The simple counterfactual test for difference-making, you will recall, was prevented from declaring influviation a difference-maker because of the following fact: if Rasputin's attackers had not drowned him they would certainly have killed him in some other way. The kairetic criterion has no such problems; for Rasputin's influviation to qualify as a kairetic difference-maker, it is sufficient that there exists a single veridical model for death from which the influviation cannot be removed. There is such a model, namely, a model that makes no mention of Rasputin's attackers' determination to kill him. Of course, there exist other veridical models for death that do cite the assassins' resolution, and from which influviation can therefore be removed without invalidating the entailment of death. But these models cannot compromise the status of influviation as difference-maker: one model is enough to confer difference-making power irrevocably. Alternative models, you will see, can never subtract from, but only add to, the list of difference-makers. In particular, Rasputin's attackers' murderous resolve will be counted as a difference-maker, but alongside, rather than in place of, his influviation.

The decisive difference in this case between the kairetic and the counterfactual accounts is, then, that whereas on the counterfactual account, an event *c*'s

making a difference to *e* depends entirely on its playing an essential role in a certain uniquely significant veridical model for *e* (roughly, a model specifying the complete state of the world up to the time that *c* obtains), on the kairetic account, *c* is a difference-maker if it plays an essential role in any one of a wide range of veridical models. The kairetic criterion for difference-making is therefore rather weaker (though not strictly weaker) than the counterfactual criterion, which in the Rasputin case and others involving "backup causes" is just what is needed. Some more challenging scenarios involving backup causes will be examined in section 6.2.

3.45 A Disjunctive Twist

Let me defend the eliminative procedure, and more broadly, the entire kairetic approach, from a familiar objection. The eliminative procedure's appeal to facts about what does and does not play an essential role in an entailment calls to mind a similar appeal by Hempel's DN account of explanation, which requires that at least one law be essentially involved in the deduction of the explanandum, and various attempts to make the hypothetico-deductive account of confirmation more sophisticated, which hold that only those parts of a theory which play an essential role in making a particular prediction are confirmed by that prediction. There is a deep problem with all these appeals to facts about what is and is not essential to an entailment. In what follows I pose the problem for the kairetic account of difference-making.[3]

Begin with a veridical, deterministic model for Rasputin's death that mentions the gravitational influence of Mars. Take some other element of the model that incontrovertibly plays a part in entailing death, say, the proposition *d* that Rasputin was thrown into the river. Replace *d* with the proposition $c \supset d$, where *c* describes the influence of Mars. This does not affect the model's

3. Concerning the DN account of explanation and the problems arising from the appeal to essential roles in entailment, see Salmon (1990b), §1.1. For a discussion of similar difficulties for hypothetico-deductivism, see Glymour (1980, chap. 2), and for Mackie's account of causation, Kim (1971).

entailment of the death, nor, apparently, the model's veridicality. Yet *c* cannot be removed from this new model without destroying the entailment of death. Thus *c*—the influence of Mars—is after all a difference-maker.

This would be a telling objection if the kairetic account required only that, in order to qualify as a difference-maker, the influence of Mars should play an essential role in entailing Rasputin's death. But more is needed: the influence of Mars must play an essential role in *causally* entailing the death.

This requirement is not satisfied. The derivation of death from Mars' influence by way of $c \supset d$ (where *c* is Mars and *d* is influviation) does not correspond to a real causal process. In particular, the step in which *d* is derived from *c* by applying modus ponens to $c \supset d$ does not correspond to the actual causal process by which *c*, or Mars, had its causal influence on the death. The actual process involved the law of gravitation; $c \supset d$ does not pick out this or any other causal law. (You will recall from section 3.23 that it is a necessary condition on causal entailment that in every application of modus ponens, the conditional should be a causal law.) Thus the entailment of death is not causal entailment, and the corresponding model not a causal model. It is therefore powerless to determine difference-makers.

I cannot emphasize strongly enough that the use of entailment in the kairetic account of explanation is quite different from its use in the DN account. The DN account, which of course epitomizes the logical empiricists' "syntactic" approach to philosophy, jettisons from the philosophy of explanation what are held to be unacceptably metaphysical notions, in particular, causal notions, substituting a logical relation, entailment, whose structure bears the entire burden of deciding what is and is not a potentially good explanation. The kairetic account, by contrast, uses entailment to represent causal relations. What is important in the kairetic account, then, is not so much the structure of the entailment relation as the causal structure of the world; above all, it is the world's causal structure that is principally responsible for determining the structure of what I am calling the causal entailment relation.

In dealing with this familiar recipe for the creation of counterexamples to syntactic philosophizing, note, I am doing something that should be utterly familiar and unobjectionable to anyone sympathetic to the causal approach to explanation: it is what all causalists do to handle the case of the flagpole and the shadow. The flagpole/shadow problem arises, recall, because the deductive argument by which the length of the shadow "explains" the height of the flagpole is identical in all relevant formal respects to the argument by which the height of the flagpole explains the length of the shadow (section 1.41). The causal approach to explanation breaks the tie by claiming that the second entailment, but not the first, represents a real causal process.

I have made the same move in this discussion: I have claimed that some entailments do, and some do not, represent causal processes, and that only the latter are suitable vehicles for causal explanation. Both the standard causal treatment of the flagpole/shadow problem and my treatment of the "disjunctive twist" assume, of course, that the facts about causal influence are rich enough to make distinctions between different entailments, but this is a sine qua non of the causal approach to explanation, not some additional demand imposed on the influence relation by my particular handling of the twist.

A different kind of attempt to deal with the disjunctive twist would blame, not the course taken by the derivation of *d* from *c* in the model's follow-through, but the presence of $c \supset d$ in the setup, on the grounds that such "disjunctive" states of affairs are for some reason inherently unfit for explanatory work. I can understand the appeal of the approach: given the form of this world's fundamental causal laws, it appears that there is simply no role for such a state of affairs to play in a causal process. Thus it can be ruled out of a causal model's setup irrespective of the form of the model's follow-through. Yet I think that this line of thought overgeneralizes: in unusual circumstances, a disjunctive state of affairs might be a condition of application for a causal law, that is, it might be the *Z* in *If F in conditions Z, then G*, for a mechanism that is set up in a very particular way. Thus it might play a valid role in a causal

entailment, genuinely representing the course of some causal process. For this reason, disjunctive states of affairs ought not to be excluded unconditionally from causal models; it is the form of the follow-through that should decide the question.

3.46 An Unhealthy Dependence on Laws?

The causal models invoked by the kairetic account appear to contain a wide range of subtle and powerful laws of nature. Consider, for example, the way that the kairetic account deals with Rasputin's poisoning (section 3.43). The poisoning is determined not to make a difference because one of the antecedent conditions in the deterministic causal law of the form

> When a person is poisoned in such and such a way, and conditions *Z* obtain, that person will die.

is not satisfied by the actual events leading up to Rasputin's death.

Fully fleshed out, this is a remarkable law statement. It is framed in terms of high level kinds such as persons, poisons, and death, yet it is not only strictly deterministic, it specifies precisely the circumstances under which poisoning is guaranteed to lead to death. Can there really be a law of nature corresponding to the statement?

Probably not. But the existence of such a law is not required for the kairetic account to do its work. A deterministic causal model for an event *e* that cites an event *c* in its setup need not cite a single law covering the entire transition from *c* to *e*; what is needed is group of laws that jointly ensure, in the circumstances, that the transition occurs. These may each cover a link in the chain from *c* to *e*, but even that is not necessary. Many different laws may work in tandem to realize a single link. In the case of Rasputin's poisoning, then, though the statement above may not pick out a single law, it will follow from other, biologically more basic, laws.

But this claim, too, might sound tendentious: surely even the basic biological laws have exceptions, and so fail to imply anything deterministic? Perhaps.

In a deterministic world, however, something like the poisoning generalization must follow from the fundamental level laws (and correspondence rules), if not from any set of higher level laws. Thus there is a basis at the fundamental level, for the assumed deterministic generalization—and this is all that the kairetic account requires in the way of laws. Indeed, if it turned out that the only genuine laws of nature were fundamental laws of nature, the kairetic account would function perfectly well, since the role of laws in the account is to determine what causally produces what, and as explained in section 3.23, this is a wholly fundamental level matter.[4]

Suppose, then, that the fundamental laws of nature, correspondence rules, and various background conditions entail the poisoning generalization stated schematically at the beginning of this section, and that it is in virtue of the generalization so entailed that Rasputin's poisoning made no difference to his death. A further problem: on the kairetic account, to know that poisoning is not a difference-maker, you must know something about the form of the "poisoning law", but how, if the generalization's provenance is some fantastically complicated entailment based in the fundamental laws, could we possibly know anything about such a law? The law is beyond us, yet the judgment that poison makes no difference, based according to the kairetic account on knowledge of the law, is easily made. Surely, then, the kairetic account, even if technically correct, cannot capture the psychology of our criterion for difference-making?

It is true that there is much about the "poisoning law" that we do not know, and perhaps will never know, if only for lack of interest. But we are quite capable of discerning the facts about the law that need to be known to make accurate difference-making judgments in many cases. Suppose that a certain poison, when it kills, does so within an hour. Rasputin is poisoned, but lives on for two hours before suffering influviation. It follows that one of the conditions required for the operation of the poisoning law did not obtain. You have no

4. I assume here a metaphysics of causal influence that draws only on fundamental laws; any of the three theories discussed in chapter one would, I think, qualify.

idea, perhaps, which one, but you do not need this information to make the judgment that poisoning was not a difference-maker for death. This, I hardly need add, is exactly the situation we are in with respect to Rasputin's actual death: no one, as far as I know, can say what saved Rasputin from the poison, but his survival shows that something saved him. The observation generalizes: it is often possible to see that one of a law's antecedent conditions did not obtain, or conversely, that all antecedent conditions did obtain, without having much idea what the antecedent conditions are.[5]

A related question: is our explanatory practice really so fixated on laws? Do we genuinely think of the process by which a certain poison has its effect as, in essence, a law-governed process?

Consider the following alternative. We think of the relevant processes as driven by underlying mechanisms. The operation of the mechanisms is based in fundamental physical processes. In a deterministic world, every mechanism has a set of enabling conditions: when the enabling conditions obtain, a mechanism runs to completion; otherwise not. In Rasputin's poisoning and like cases, what we recognize, without understanding every detail of the mechanism's operation, is that some of the enabling conditions did not obtain.

This alternative picture is, I suggest, merely a different way of expressing the same facts as the law-based picture; this correspondence is due to the extremely close relationship between fundamental laws and mechanisms, discussed in section 7.6

5. The case of the indeterministic bomb under the chair (sections 2.3 and 11.3) shows that it is possible to see that something derailed a process even when, in a certain sense, there is no fact of the matter as to what went wrong.

3.5 Abstraction and Optimizing

3.51 Elimination Is All or Nothing

The principal weakness of the eliminative procedure lies in its offering only two choices in dealing with a causal factor: either the factor is completely removed from the model in question, or it is retained in its entirety. Frequently, what is wanted is something between these two extremes, as the following example shows.

I throw a cannonball at a window, and the window breaks. The cannonball weighs exactly 10 kg. Does the fact that the cannonball weighs exactly 10 kg make a difference to the window's breaking? The natural answer to this question is no. The fact the cannonball is rather heavy made a difference, but the fact that it weighed in at exactly 10 kg did not. You would like to say, then, that a certain fact about the ball's mass—that it was rather heavy—made a difference, but that finer details about the mass did not.

The eliminative procedure does not allow such delicate distinctions. Take a model for the window's breaking that includes in its setup only one proposition concerning the cannonball's mass, namely, a proposition stating that the mass was 10 kg. If the proposition is removed from the model, there is no longer any constraint on the ball's mass at all. The new setup, then, is compatible with the ball's having the mass of a grain of sand; consequently, the setup will no longer entail the window's breaking. Because the fact of the ball's exact mass cannot be removed from the model, the exact mass is counted by the eliminative procedure as a difference-maker for the breaking, an unwanted conclusion.[6]

6. The eliminative procedure does count the fact of the ball's being rather heavy as a difference-maker; simply start with a model that says no more about the ball's mass than that it is considerable. The problem, then, is not that it is impossible to count the fact of rough mass as a difference-maker, but that the exact mass, too, counts as a difference-maker.

3.52 Abstraction

The problem is solved by introducing a way of removing detail from a causal model's setup that stops short of completely excising a proposition. You would like to take a proposition such as *The ball's mass was 10 kg* and make it less exact, so that it says, say, *The ball's mass was greater than 1 kg*. I will call the suggested operation *abstraction*.

Say that one model M is an abstraction of another model M' just in case (a) all causal influences described by M are also described by M', and (b) M' says at least as much as M, or a little more formally, every proposition in M is entailed by the propositions in M'. Intuitively, if M is an abstraction of M', then M' may be obtained by "fleshing out" M, that is, by adding some additional causal details to M's description of a causal process.[7] The abstraction ordering is a partial order in the technical sense.

Let me construct a replacement for the eliminative procedure that substitutes abstraction for removal. Whereas the eliminative procedure transforms a causal model M for an event e into a kernel for e by removing elements of M until no further elements can be removed without invalidating M's entailment of e, the new procedure transforms M into a kernel by performing abstraction operations on M's setup until no further abstraction is possible without invalidating M's entailment of e.

As an illustration, apply this abstraction procedure to the case of the cannonball's breaking the window. Suppose you begin with a causal model for the breaking that contains a statement of the cannonball's exact mass. The eliminative procedure could not touch this statement; its removal would invalidate

7. Why is part (a) of the definition necessary? There are two senses in which M might be more abstract than M': it might say less about the causal process represented by M' than M' itself, or it might say less about the causal influences represented by M' than M' itself. It is the second, strictly stronger, notion that is wanted here. What is the difference? Suppose that M' tells the complete story about the deterministic causal production of e starting two minutes before e occurred. A model that tells the complete causal story starting one minute before e occurred is an abstraction of M' in the first sense (because its entire setup is entailed by M''s setup), but not the second.

the entailment of the breaking. The new procedure can, by contrast, make the claim more abstract, that is, less detailed, resulting in a setup that, rather than stating an exact mass for the ball, states a range of possible masses. The setup contains, in other words, a statement of the form *The ball's mass was between a and b kg*. (The range must include the actual mass, of course, or the statement would be false and the setup non-veridical. Because any abstraction of a veridical model is itself veridical, this requirement need not be made explicit.)

Suppose that any mass for the ball greater than 1 kg is sufficient, in conjunction with the other states of affairs stipulated by the setup, to entail that the window breaks. Then the maximum abstraction possible without invalidating the entailment of the breaking would result in the setup's saying *The ball's mass was greater than 1 kg*. Consequently, a model for the breaking that involves the ball's mass will, upon application of the abstraction procedure, end up making just this statement about the mass. Thus you have the conclusion you want: the one and only fact about the mass that is a difference-maker for the breaking is the fact of the mass's being greater than 1 kg. You can say that it mattered that the ball was heavy, but it did not matter how heavy it was.

A cannonball's breaking a window is not an event of great scientific interest, but the ability to abstract rather than merely to remove the elements of a causal model in the search for explanations is important to science as well as to common sense. I will argue in later chapters that abstraction underlies the many important explanatory phenomena already mentioned in the preface: explanatory omissions, equilibrium explanations, robustness, idealization, the use of probabilistic explanation in deterministic systems, and so on.

The move from elimination to abstraction clarifies, by the way, an earlier claim about the kairetic account's treatment of causal factors such as Mars' gravitational influence on Rasputin's death. When treating this case in section 3.42, I asserted that the eliminative procedure deleted from a model for Rasputin's death any part of the setup stating the magnitude of Mars' gravitational pull, while "leaving behind" a condition specifying that there were no significant

gravitational influences on the process except for Earth's. If such a negative condition were already present in the model's setup then things would work as stated; you can now see that even if the condition were not present, it would be created by abstraction from the specifications of various gravitational fields of non-terrestrial origin. The model, in other words, would summarize the explanatorily relevant information about these fields simply by saying that they were not very large.

All elements of a causal model are subject to abstraction: not just its events and other singular states of affairs, but also its laws. Consider, for example, window breaking. A causal model for a cannonball's breaking a window might contain many details about the laws of physics, such as the complete theory of gravity, the physics of air resistance, and the laws governing the molecular structure of the window. Much of this detail can be removed. The laws of pertaining to the cannonball's trajectory can be abstracted to the simple law discovered by Galileo governing the motion of terrestrial projectiles. The laws concerning molecular detail can be abstracted to a simple law about the brittleness of ordinary glass. An explanatory kernel, then, will cite only these expurgated versions of the underlying laws.

Or take Rasputin's death. A causal model for Rasputin's death by drowning might contain many complex physiological generalizations about the precise effects of oxygen deprivation, among other things. If the explanandum is simply death, however, the precise effects are irrelevant, provided that they result in Rasputin's dying. They may be abstracted away, then, leaving behind the simple relation between oxygen deprivation and death. There are limits to this abstraction, but I will not dwell on them here; they will be explored in sections 3.6 and 5.4.

With some of their details removed by explanatory abstraction, is it right to say that the abstracted laws are nevertheless difference-makers for the explanandum? They are not difference-makers in their entirety; rather, certain aspects or high level properties of the laws are difference-makers. These aspects

take the form of more abstract versions of the original laws; they are themselves law-like. The process of kernel determination thus takes the raw nomological material present in the source model and produces its own high level laws, each with a level of abstraction tailor-made for the explanatory task at hand, so as to specify only properties and processes that are relevant to the explanandum.

This is the case even when the source model describes a causal process entirely at the level of fundamental physics. The high level laws in the resulting kernel should be regarded as nothing new, then, but merely as a more abstract description of the same fundamental level causal process, or in other words, as high level descriptions of the fundamental level laws (and perhaps other fundamental level states of affairs) obtained by omitting all non-difference-making details in the fundamental level model. In chapter seven I will propose that all high level causal laws, not just those produced by kernel determination, have what I call "underlying mechanisms" that are simply the kairetic procedure's distillations of the fundamental laws. In this way, the kairetic procedure is in a certain sense responsible for the structure of the high level sciences.[8]

3.53 Path Dependence and the Optimizing Procedure

The procedure for determining an explanatory kernel by abstraction requires some further tuning, as a slightly more complex cannonball/window example will show. Suppose that, in order to break the window, the cannonball must have a momentum of at least 20 $kgms^{-1}$. (The momentum of the cannonball is its mass multiplied by its velocity.)

The actual cannonball had, say, a mass of 10 kg and a velocity of 5 ms^{-1}. You would like to know what facts about the ball's mass and velocity were difference-makers for the breaking. You ought to be able to answer this question by taking a causal model for the breaking that states the exact mass and velocity,

8. There is one other principal ingredient of the high level laws, and thus one other important determinant of high level scientific structure: the basing generalizations discussed in chapter seven.

and then abstracting until you can abstract no further without invalidating the entailment of the breaking. Whatever claim about mass and velocity remains spells out the difference-making facts.

In implementing this procedure, a problem arises: the kernel determined by the procedure, and in particular the claims about the mass and velocity of the ball that appear in the kernel's setup, will vary depending on the order in which the abstractions are performed. The procedure, then, does not provide a well-defined set of difference-makers.

Suppose, for example, you begin with a model that states an exact mass and velocity for the ball, namely, 10 kg and 5 ms^{-1}, and you decide to abstract velocity, that is, to substitute for the specification of the exact velocity a specification of a range of velocities. The widest, hence most abstract, range you can specify without invalidating the entailment of the window's breaking is that the velocity is 2 ms^{-1} or more. If the lower bound on the velocity is reduced any further, the momentum of the ball will, given that its mass is specified to be 10 kg, fall below 20 $kgms^{-1}$, and so will no longer be sufficient to entail that the window breaks. Now abstract the mass. Because the lower bound on velocity is 2 ms^{-1}, the mass cannot be reduced at all; you are therefore left with a specification that the mass is 10 kg or greater. In summary, the difference-making facts about mass and velocity turn out to be: the mass was equal to or greater than 10 kg, and the velocity was equal to or greater than 2 ms^{-1}.

But now suppose that you had abstracted the mass first. With the velocity fixed at this initial stage at its actual value of 5 ms^{-1}, you can specify a range for the mass of 4 kg and up. When you come to abstract the velocity, you are left with a range of 5 ms^{-1} and up. Thus the difference-making facts about mass and velocity turn out to be: the mass was equal to or greater than 4 kg, and the velocity was equal to or greater than 5 ms^{-1}, in conflict with the conclusion of the previous paragraph. You will see that there are other possible end points of the abstraction process, too: for example, you might conclude that the difference-making facts were that the mass was equal to or greater than 5 kg and

that the velocity was equal to or greater than 4 ms^{-1}.

The conflict is resolved by specifying a unique end point for the abstraction operation, as follows. Given a model M for an event e, the corresponding kernel for e is the maximal abstraction of M that causally entails e. Because the abstraction relation is only a partial ordering, it is not guaranteed that M will have a maximal abstraction, that is, an abstraction that is more abstract than any other abstraction. But in the case of the cannonball, there is such a model. It specifies the following constraint on the mass m and velocity v of the ball:

$$mv \geq 20.$$

You will see that a kernel built around this specification is at least as abstract as any of the possible end points considered above. It is also clearly the most abstract such specification: a more abstract specification would have to allow a momentum for the ball less than 20 $kgms^{-1}$, and in so doing, could no longer guarantee the breaking of the window.

What if there is no unique maximally abstract model? For now, let me say that in such cases there are no determinate facts about difference-making, or rather, no determinate facts save those on which all the most abstract models agree.

Whereas the earlier procedures for kernel determination specified a process, the new procedure specifies a goal, the maximally abstract model. I call this new procedure the *optimizing procedure*—or at least, I will do so once the problem of cohesion is addressed.

3.6 Cohesion

3.61 A Disjunction Problem

A disjunction always says less than its disjuncts. Thus it is always possible to make a causal model more abstract by forming a disjunction of the model's setup and the setup for some other causal model with the same target, resulting

in a set of uselessly disjunctive difference-makers. This spells trouble for the optimizing procedure. Let me give an example.

Consider two causal models that represent Rasputin's dying in two quite different ways, say, a death by drowning due to his being thrown into a river, and a death by poisoning due to his being fed toxic teacakes. Suppose that the influviation model contains the real difference-makers for Rasputin's death; the poisoning model is therefore not veridical.

Now take the disjunction of the setups of the two models, and form a new model that has the disjunction as its setup: it states that *either* Rasputin was thrown in a river etc., *or* he was fed poison teacakes, etc. The disjunctive model is veridical, since one of these chains of events did occur, as claimed, and it entails Rasputin's death, since both chains of events lead to death. Worse, it is more abstract than the influviation model, because its setup is entailed by the influviation model's setup but not vice-versa.

It would appear, then, that the optimizing criterion for difference-making will favor the disjunctive model over the influviation model. As a consequence, Rasputin's being thrown into the river will not qualify as a difference-maker for his death; the difference-maker is rather the more abstract event of his being either thrown into the river or poisoned. (More exactly, the difference-maker is the state of affairs picked out by the disjunction of the descriptions of the two causal chains in their entirety.) This is not a tolerable conclusion. Even if the disjunction can be said in some extenuated sense to be a difference-maker, the disjunct—the influviation—ought to be a difference-maker too.

3.62 The Cohesion Requirement and Its Foundations

The disjunctive model is defective, but in what way? It is tempting to say that a model with a single premise that lumps together events, laws, and background conditions in an inextricable way is not a genuine causal model, perhaps because its entailment of its target is not causal entailment. There may well be something to this, but I am unwilling put too much more weight on the notion

of causal entailment, and in any case, there is an opportunity here to pursue a deeper issue.

Rather than the form of the disjunctive model, focus on the different kinds of scenarios that can realize the model. These fall into two classes: the poisonings of Rasputin, and the influviations of Rasputin. There is something wrong, I suggest, with a causal model that is realized by, and only by, two quite different kinds of causal processes. I say—alluding to the property of the same name employed by the pattern subsumption approach to explanation (section 1.32)—that such a model lacks *cohesion*. Even after application of the kairetic criterion has eliminated such a model's non-difference-makers, it is unable to function as an explanation because it models not one but two distinct difference-making processes. To use such a model to pick out the difference-makers in a causal process cannot succeed, because to say that the process realizes the model leaves it indeterminate which difference-making process is at work. This accounts for the unsatisfying indeterminacy of the states of affairs deemed difference-makers by a disjunctive kernel. I therefore make the following amendment to the optimizing procedure for kernel determination: an explanatory kernel must be cohesive. The cohesion requirement acts as a brake on abstraction, halting it before it crosses the line into disjunction. In so doing, the requirement not only prevents technical philosophical trickery such as the disjunction problem, but also determines how abstract a description of the workings of the world can be while still giving explanatorily useful information about fundamental level causation.

I have given only the loosest and most intuitive characterization of cohesion. What is it for a model to be realized by two different causal processes? In Strevens (2004), I defined cohesion as follows. The disjunctive model for Rasputin's death mentions a number of causal elements: teacakes, toxins, ropes, rivers, and so on. These can be divided into two sets—the sets on opposite sides of the disjunction—corresponding to the two ways that the model can be realized. Some systems realize the model by possessing the elements in one set,

some systems by possessing the elements in the other. Call a model incohesive, then, to the extent that its different realizers appeal to different sets of the causal elements mentioned in the model, or in other words, call a model cohesive only if all of its realizers possess the same causal elements.

This way of defining cohesion in effect explains what it is for a model to describe two different causal processes by appealing to a distinction between different causal elements: causal processes are individuated by their elements. But how are causal elements to be individuated?

Let me consider two possibilities. First, there is perhaps some metaphysical or psychological criterion for individuating causal elements that lies entirely outside the sphere of explanation, and to which philosophers of explanation may appeal without further justification. Suspecting as I do that high level causal talk is in part explanatory talk, I do not regard this as a secure option.

Second, differences between causal elements are perhaps to be discerned at the level of fundamental physics. Certainly, fundamental physics is capable of recognizing any difference you like, but there may be too much discernment and so too many differences: every concrete realization of Rasputin's death looks somewhat different from every other.

There is something of a dilemma in the making, then. If the causal difference between a poisoning and a drowning is something that emerges only at the high level, then like all high level causal facts, on my view, it is in part an explanatory fact and so should be given a basis in a theory of explanation, rather than forming a part of the basis for a theory of explanation. This suggests a fundamental level basis for the distinction I need. But from a fundamental level perspective, every causal process is unique. To forbid causal models that can be realized by different causal processes would be to forbid abstract models altogether.

What to do?

3.63 Cohesion as Causal Contiguity

My solution is to opt for a fundamental level basis and to invoke a notion of contiguity. Suppose that fundamental level causal processes can be rated as more or less similar to one another on a continuous scale of similarity, so that all such processes form a similarity space. Any particular realization of a causal model will correspond to a point in this similarity space. A model is cohesive, I propose, if its realizers constitute a contiguous set in causal similarity space, or as I will say, when its realizers are *causally contiguous*. Thus two realizers of a causal model may be quite dissimilar without disrupting its cohesion, provided that it is possible to trace a path through causal similarity space from one realizer to the other without passing through any non-realizers.

In its broad outlines, this definition of cohesion solves the problems outlined in the previous section: it gives cohesion a fundamental level basis, but admits models with realizers that are from a fundamental level point of view (or even a high level point of view) extremely diverse.

To posit a similarity space subsuming every kind of causal process may seem rather optimistic, however. Let me remind you what is at issue here. The ultimate goal is characterize our practice of scientific explanation. To base cohesion on causal contiguity, then, is to assume that our practice includes procedures for making similarity judgments about causal processes and that these judgments, if they exist, play a certain role in evaluating explanatory models. I take it that it is the existence of the judgments themselves that is more likely to raise doubts.

Two remarks. First, it is not necessary to show that the similarity judgments are based on some objectively real property of resemblance. If the similarity relation in question is a part of our explanatory practice, then it belongs in an account of scientific explanation, even if it is a mere artifact of human cognition. Second, it is not necessary to show that, given any two causal processes, a scientific practitioner can assess their degree of similarity. What is needed is that, given a causal process, the practitioner can identify its close

neighbors in the similarity space, that is, that the practitioner can identify small "similarity neighborhoods".[9]

Nevertheless, the obstacles to finding an account of similarity judgments are considerable. Even if a cognitive basis for the judgments does exist, it may be a complex psychological matter that is for the most part beyond the means of contemporary philosophy to discover. How even to begin creating a formal system capable of representing all possible causal processes, especially when *possible* means, as you will see, not physically possible but metaphysically possible?

For the purposes of this study, I propose a proxy for causal contiguity that I call *dynamic contiguity*. Dynamic contiguity is a necessary condition for causal contiguity, I will suggest, and in certain circumstances, a sufficient condition. Thus it can be used to flesh out the notion of cohesion—and as a necessary condition, to provide decisive judgments about incohesion—without my having to investigate further the nature of the similarity of causal processes.

Every causal process—if you like, every concrete realization of a causal model—corresponds to a trajectory in the state space of fundamental physics. (State space contains a point for every possible state of a given system. A trajectory in state space therefore traces the way in which the system's properties change as a causal process unfolds.) Take the trajectories corresponding to every one of a model's concrete realizers, each a thread in state space. If this set of state space threads is contiguous, the model is *dynamically contiguous*.

Were it the case that a trajectory in state space represented all the facts about the corresponding causal process, then causal contiguity might be defined as dynamic contiguity. But a trajectory contains information about causation only indirectly; it reveals the overt effects of the causal influences at work in a system, but it does not reveal the nature of the influences themselves. The

9. To put it technically, contiguity requires for its foundation only a topology, not a metric; in practice, I would guess, the foundation for the topology is likely to be a local metric, and judgments of long-distance similarity, though unnecessary, will not be unknown.

same trajectory in state space might be caused in two quite different ways, corresponding to two different sets of possible fundamental physical laws. (If the fundamental laws are fixed, then there is, given determinism, just one causal process per trajectory, since "the system" in a fundamental level causal model by definition comprehends all causal influences. But as you will see, it is important to the kairetic account that cohesion be defined for models that have realizers with fundamental laws different from the actual fundamental laws.)

I propose that dynamic contiguity is necessary for causal contiguity, at least when working with the kinds of physical theories that seem to characterize our universe. A model with dynamically non-contiguous realizers, then, is not cohesive. Dynamic contiguity conjoined with a certain additional factor is sufficient for causal contiguity. What is the extra ingredient? I have only a sketch of an answer: the dynamic contiguity of a model guarantees its causal contiguity provided that either (a) the model allows only one fundamental physics, that is, all concrete realizations of the model agree on the fundamental laws, or (b) the different, competing sets of fundamental laws allowed by the model themselves form a contiguous set in some appropriate similarity space for fundamental physical theories. In this way, the problem of finding a causal similarity space is reduced to the problem of finding a dynamic similarity space along with a similarity space for fundamental theories. I would like to think that this constitutes progress.

In any case, although dynamic contiguity alone is not sufficient for causal contiguity, hence not sufficient for cohesion, I propose nevertheless that, within the framework of the kairetic criterion for difference-making, the dynamic contiguity test can for the most part be safely used in place of the causal contiguity test. The shortcomings of dynamic contiguity as a mark of cohesion will, I suggest, be covered up in almost all cases by the kairetic criterion's generality desideratum: models that are wrongly judged cohesive by the dynamic contiguity test will be rejected anyway for being insufficiently general. Why? When the sets of fundamental laws cited in the various concrete realizations of a model are

not contiguous there will exist, I am guessing, a more abstract model that fills in the gaps, and so (given dynamic contiguity) passes the test for cohesion. But I will not try to justify this claim here. Let me simply take dynamic contiguity as a proxy for cohesion and let the results speak for themselves.

An example: consider the trajectories corresponding to the concrete realizations of a causal model in which Rasputin dies by influviation. These, I submit, form a contiguous region in the state space of fundamental physics. That is, you can get from one realizer to any other by making (speaking roughly here) a perhaps very long series of very small changes in the realization, adjusting incrementally the place where Rasputin hits the water, the temperature of the water, the weight and composition of his clothes, and so on, in such a way that every intermediate stage is itself a realizer of the model. The influviation model therefore passes the dynamic contiguity test for cohesion. The trajectories corresponding to the realizers of the disjunctive model do not, by contrast, form a contiguous set.[10] Thus the disjunctive model is, as desired, judged incohesive.

Three remarks. First, how to compare realizers with different numbers of particles? The corresponding trajectories occupy differently dimensioned state spaces, the smaller of which cannot be meaningfully embedded in the larger. A similarity metric that reaches beyond any particular state space is required, then, to determine which n particle systems are more, which less, similar to a given $n + 1$ particle system.

Second, as I have already observed, the test for cohesion must deliver judgments about models that have physically impossible systems among their realizers, that is, models that are consistent with sets of fundamental laws other than the actual laws of fundamental physics. In chapter eight, for example, I will consider a model that can be realized by both classical and quantum

10. Or if they do—by way of a simultaneous poisoning/influviation where an infinitesimal adjustment makes the difference between death by poison and death by drowning—they touch only at a singularity, and so are not contiguous in a more robust sense (see below).

systems. Classical and quantum physics have rather different state spaces. If the dynamic contiguity approach is to succeed in evaluating the cohesion of such models, there must be some way of projecting trajectories in these different spaces onto some common space where their contiguity can be evaluated. In particular, there must be some way of deciding whether a classical and a quantum theory are making almost the same or quite different predictions about the time evolution of a given system.

To find a basis for such judgments is a difficult problem. Yet it is clear, I think, that we are in fact capable of comparing predictions in this way, at least in some cases. We can, for example, compare the predictions of classical and quantum theories in situations where, on the quantum side, well-defined wave packets are involved. (For another example, see section 8.24.)

Third, it is quite possible that, on top of connectedness in the technical sense, a set of trajectories should satisfy some further geometrical constraints if it is to qualify as cohesive. I have already suggested (in note 10) that all parts of the set ought to be of the same dimension, and in particular, that the contiguity of the set ought not to be due to low-dimensional "bridges" between its parts. Even trajectory sets that satisfy this additional requirement may, if their geometry is too extreme, fail to measure up to our intuitive standards for cohesion in some other way.

The notion of dynamic contiguity needs further work, then. I am in any case not entirely confident that causal contiguity is the key to understanding what is explanatorily unsatisfying about disjunctive models. A contiguity criterion for cohesion does have certain advantages, however: it is a fundamental level criterion, so does not lean on a metaphysics of the higher level, and it is a strong principle, whose consequences for the nature of explanation will be interesting and controversial. Let me proceed, then, on the understanding that some fundamental level account of cohesion is required, and that the contiguity account is as good a placeholder for the correct account as any.

An aside: In endorsing the contiguity conception of cohesion, I am claiming

that our explanatory practice is committed, in principle, to giving explanations by way of models that are realized contiguously at the fundamental level. It does not follow that every explanation is explicitly vetted for this property, or even that human explainers have much idea what the fundamental level looks like. Cohesion may turn in principle on the nature of the fundamental level, but in practice, the cohesiveness of an explanatory model is likely to be tested by high level heuristics. Explanatory claims, like all other scientific conclusions, are forever—well, almost forever—provisional.

Let me examine, briefly, two controversial consequences of the contiguity conception of cohesion. Consider the cohesion of a model that attributes Rasputin's death to poisoned teacakes. Different poisons act in different ways. It seems likely, then, that the various realizations of poisoning will not form a contiguous set at the fundamental level. The poisoning model is therefore incohesive: it should be broken down into cohesive models each representing the effect of a certain kind of poison on Rasputin's system. Strictly speaking, if Rasputin had (contrary to the facts) died from poisoning, it would not have been poisoning per se that made a difference to his death, but a certain kind of poisoning.

Is this consequence tolerable? I think it is not only tolerable but has the ring of truth. What makes something a poison is (in part) that it causes death. But to say that Rasputin was killed by ingesting something that causes death is to give an incomplete explanation of the death; a model that cites poison without saying more is therefore explanatorily defective, just as the kairetic criterion would have it. The kairetic account's treatment of multiply realizable kinds is discussed further in section 5.4; on functionally defined kinds in particular, see section 4.33.

Next consider the model that attributes Rasputin's death to influviation. Suppose that, as a matter of definition, influviation involves the victim's being bound hand and foot. Being bound is surely a multiply realizable property: it can be done with rope, wire, duct tape, and so on. Do influviations performed

using these various kinds of binding form a dynamically contiguous set? It is hard to say. There are discontinuities between different molecular structures, but the points of configuration space between the structures do not represent equilibrium states. However, it is implausible that the explanatory power of the influviation model depends, or depends that much, on the resolution of such questions. (Compare my second criticism of the counterfactual criterion for difference-making in section 2.4.)

Two responses. First, it may be that the answer to the question about binding makes little or no difference to most explanatory questions because, even if there were a separate explanatory model for every mode of binding, these models would, for the most part, identify the same difference-makers for death: the water, the sinking, and so on. Thus you can extract determinate facts about difference-making even if you do not know enough about the physics of binding to determine the correct kernel: all plausible candidates for the kernel concur in almost all their judgments of difference-making.

Second, if cohesion is made a desideratum rather than a requirement, then the influviation model can be seen as one that trades a relatively small amount of cohesion for a big increase in abstractness. It does not maximize cohesion, but it maximizes combined cohesion and abstractness. Such tradeoffs are discussed further in section 5.43.

3.7 The Optimizing Procedure

According to the optimizing procedure for kernel determination, the explanatory kernel corresponding to a veridical deterministic causal model M with target e is the causal model K for e that satisfies the following conditions:

1. K is an abstraction of M,
2. K causally entails e,

and that, within these constraints, maximizes the following desiderata:

3. Generality: *K* is as abstract as can be, and

4. Cohesion: The fundamental level realizers of *K* form a causally contiguous set.

I have made cohesion a desideratum rather than a requirement, as suggested at the end of section 3.63; you might also treat cohesion as non-negotiable.

Three remarks. First, condition (2) is strictly speaking redundant, as it merely reiterates the specification that *K* be a causal model for *e*.

Second, the optimizing procedure's desideratum of generality is subtly different from the unification account's desideratum of the same name: whereas the pattern subsumption account's generality requirement counts only actual phenomena subsumed, the kairetic account's generality requirement in effect counts possible but non-actual systems. (Weisberg (2003) provides an extended discussion of the significance of these two kinds of generality.)

Third, generality and cohesion should not be regarded as distinct explanatory virtues. They have no value in themselves (though see section 4.35); their sole role is the determination of difference-makers.

3.8 Rival Accounts of Difference-Making Reconsidered

One way to establish the kairetic difference-making criterion's superiority over its rivals is to display the rivals' difficulties and shortcomings, as I did in chapter two. Another, perhaps better, way is to explain the rivals' successes—to show, given the truth of the kairetic account, why they work when they do.

3.81 The Counterfactual Criterion

In singular event explanation at least, the simple counterfactual criterion is almost never wrong when it rules that an event is a difference-maker for another event. It is also usually right when it rules that an event is not a difference-maker for another event, the main exceptions being cases of preemption. Let

me explain why the simple counterfactual account succeeds when it does, and why it fails when it does, taking as my premise the correctness of the kairetic criterion. (I leave the corresponding explanation for more sophisticated variants of the counterfactual account as an exercise for the reader.)

First, the counterfactual criterion's positive rulings about difference-making. According to the kairetic criterion (at first I will employ the eliminative version), a necessary and sufficient condition for an event *c*'s making a difference to the occurrence of an event *e* is:

> There exists a model, obtained by removing *c* from a veridical deterministic causal model for *e*, that does not causally entail *e*.

The counterfactual criterion for difference-making tests whether an event *c* is a difference-maker for an event *e* by looking to the closest possible world *w* where *c* does not occur (in what follows I assume for simplicity's sake that there is always a determinately closest world). If *e* does not occur in *w*, then *c* is a difference-maker; if *e* occurs in *w*, then *c* is not a difference-maker.

Suppose the counterfactual criterion declares that *c* is a difference-maker for *e* in virtue of such a *w*. This judgment is vindicated by the kairetic criterion, provided that *w* realizes a model *M* obtained by removing *c* from a veridical deterministic model for *e*, since if *e* does not obtain in *w*, the setup of *M* cannot causally entail *e*. (In effect, *w* is simply one way of filling out the incomplete causal story told by *M*.)

The question, then, is whether the setup of some relevant *M* is true in *w*. When *c* is a particular event or local state of affairs, as I have been assuming, this is likely to be the case for any such *M*. Why? The closest world is one that differs as little as possible from the actual world before the occurrence of *c*, typically the result of "small miracle" that prevents the occurrence of *c* while changing little else (section 2.4). Whatever states of affairs are specified in *M*'s setup are therefore very likely present in *w*. The counterfactual criterion's positive difference-making judgments are for this reason generally correct (though I will qualify this claim shortly).

Next, the counterfactual criterion's negative difference-making rulings. According to the kairetic account, an event *c* fails to make a difference to an event *e* just in case every model *M*, obtained as above by removing *c* from a veridical deterministic model for *e*, causally entails *e*. The counterfactual criterion for difference-making holds that *c* is not a difference-maker just in case *e* occurs in the closest possible world in which *c* does not occur. Suppose that *c* fails the counterfactual test: in the closest world *w* in which *c* does not occur, *e* still occurs. Since *c* is a local event or state of affairs, the miracle that ensures that *c* does not obtain in *w* is very likely small enough that all aspects of the actual world that were causal influences on *e*, apart from *c* and its causal consequences, are true in *w*. Thus, as before, the elements in the setup of any relevant *c*-less model *M* for *e* hold in *w*.

Assume that *e* holds in *w* because it is causally entailed by certain facts in *w* (as best I can tell, this is usually true). It follows that for every relevant *M*, there is a set of causal factors and laws consistent with the setup of *M* that causally entails *e*, namely, the facts in *w* that causally entail *e*. But this does not imply that such an *M* itself causally entails *e*, since the causal entailers of *e* in *w* need not be represented in the setup of *M*: *e*'s obtaining in *w* may be due to some causal process other than the one represented by *M*. This is just the situation found in cases of preemption (sections 2.4 and 6.2); the kairetic account predicts, then, that the counterfactual account's negative rulings are unreliable in the presence of a backup cause.

Now, a complication: I assumed above that the correct answers about difference-making are given by the kairetic account's eliminative procedure. Thus I have established only that the counterfactual criterion's judgments of difference-making are as reliable as the eliminative procedure's judgments. It should come as no surprise, then, that when the eliminative procedure fails, the counterfactual criterion also fails.

Consider the case of the cannonball and the window from section 3.53. You will recall that a cannonball of mass 10 kg hurled at 5 ms^{-1} breaks a window, any

momentum of 20 $kgms^{-1}$ being sufficient to cause the breaking. The kairetic kernel for the breaking specifies of the mass and velocity only that their product is greater than or equal to 20.

Put the following question to the counterfactual criterion: did it make a difference that the velocity of the ball was at least 2 ms^{-1}? To answer the question, find the nearest possible world where the velocity is less than 2 ms^{-1}. Because this world is chosen to be maximally similar to our own, I assume that the ball will have a velocity barely under 2 ms^{-1} and the mass will be the same, 10 kg. The ball's momentum in this world is just under 20 $kgms^{-1}$, so the window is not broken. The ball's having a velocity of at least 2 ms^{-1} is therefore declared a difference-maker.

If the kairetic criterion is correct, this claim is false. There are realizations of the optimal model for the window's breaking in which the velocity of the ball is less than 2 ms^{-1}, for example, a realization in which the velocity is 1.5 ms^{-1} and the mass is 15 kg. The counterfactual account fails to capture this subtlety in our reasoning about difference-making.

In summary, the kairetic account of explanatory relevance is able to account for the major strengths and weaknesses of the simple counterfactual account: the account works badly in cases of preemption, but works well in other cases of event explanation, provided that the putative difference-maker is itself a singular event and so can be removed using a small miracle, and that the case is not one in which abstraction, rather than removal, is appropriate. When the putative difference-maker is not an event and the removal miracle is not so small, the counterfactual account may fail to give a clear answer, as shown in section 2.4.

3.82 The Manipulation Criterion

The manipulation criterion for explanatory relevance succeeds for more or less the same reasons as the counterfactual criterion. Like the counterfactual and kairetic criteria, to test whether an event *c* makes a difference to an event *e*, the

manipulation criterion asks whether *e* occurs in a scenario from which *c* has been "removed". Like the counterfactual criterion, the manipulation criterion considers a scenario in which *c* determinately does not occur; like the kairetic criterion, it conceives of the scenario abstractly, by way of a less than complete description of the relevant goings-on.

When the manipulation criterion makes a positive relevance judgment (with respect to a given causal pathway), it does so because *e* does not occur in a certain causal model in which *c* does not occur. Given the manipulationist's rules for determining the nature of the model, it is more or less guaranteed that the model is a fleshing out of a veridical deterministic model for *e* from which *c* has been removed in the kairetic sense—again, call it *M*. Since a fleshing out of *M* is consistent with the non-occurrence of *e*, it follows that *M* itself cannot entail *e*, and so that *c* is a difference-maker. The manipulation criterion's positive judgments of difference-making are, therefore, accurate. (However, the manipulation criterion shares with the counterfactual criterion an inability to deliver subtle judgments about difference-making in cases where the optimizing procedure yields a different model than the eliminative procedure, as explained in the previous section.)

The negative judgments made by the criterion are more accurate than the simple counterfactual criterion's negative judgments, which fall through in cases of preemption, you will recall, because the counterfactual criterion's *c*-less scenario—the scenario it examines to ascertain the consequences of removing *c*—includes too much. The manipulation criterion's *c*-less model will contain much less than is contained in a possible world. In this respect it is already ahead of the counterfactual criterion. But as the example discussed in section 2.5 and pictured in figure 2.1 shows, the manipulationist model may nevertheless contain backup causes.

The reliability of the manipulation criterion in preemption cases is due, then, to something else—namely, what it does with the backup causes when creating its *c*-less model, which is to assign every causal variable not on the

putative causal path from *c* to *e* its actual value. Schematically, a typical backup cause has two states, active and passive. It starts out in its passive state and assumes its active state, in which it is able to cause *e*, only if called on to do so by the failure of the main cause. In a standard preemption scenario, the backup cause does not lie on the causal path from *c* to *e*, and remains in its passive state at all relevant times, since the main cause is successful in bringing about *e*. The manipulation criterion, then, holds the backup cause in its actual state—the passive state—even in a *c*-less model, and so *e* does not occur in such a model. The criterion in effect removes not only *c*, but also the consequences of the triggering mechanism by which the non-occurrence of *c* would activate the backup cause.

From a kairetic perspective, then, in cases of preemption the manipulation criterion does something wrong, but also does something else that neutralizes its wrong-doing. When removing *c*, it retains the backup mechanism. (The kairetic account, by contrast, attributes difference-making power to a preempting cause *c* in virtue of a model that does not mention the backup mechanism; see section 6.2.) But it disables the mechanism, achieving much the same result as if it had entirely removed it.

Is the manipulation criterion infallible in its negative judgments of relevance? Only if it always succeeds, in cases of preemption, in neutralizing the backup mechanism. From the justification of its success above, one potential weak point is clear: since only mechanisms not on the putative causal path from *c* to *e* are disabled, the manipulation criterion will run into a certain amount of difficulty when the variables that constitute the backup mechanism themselves lie on the path from *c* to *e*—in which case the same variables play a dual role as both backup mechanism and assistant to the main mechanism, the mechanism by which *c* actually did cause *e*. Such cases are discussed in Strevens (in press a); I give an additional argument against the manipulation criterion in section 6.25.

3.83 The Mill/Mackie Criterion

I neglected in chapter two the influential criterion that, following J. S. Mill, looks for necessary and/or sufficient conditions for the explanandum. This approach is best captured by Mackie (1974)'s notion of an INUS condition. As with most other philosophical attempts to articulate a notion of difference-making, Mackie's criterion was presented as the core element of an account of high level causation. (For an account of explanation with hints of Mackie, see Garfinkel (1981), especially pp. 62–66.) I reinterpret Mackie's theory as a criterion for difference-making, as follows.

According to Mackie, *c* is a difference-maker for *e* if *c* is an insufficient but non-redundant part of an unnecessary but sufficient condition for the occurrence of *e*. That is, to find a cause of *e*, find some bundle of conditions that is sufficient for *e*—such as the conditions asserted by a deterministic causal model for *e*—and pick out a non-redundant part of the bundle. What is a non-redundant part? It is a part of the bundle that cannot be removed without making the bundle no longer sufficient for *e*.

You will see that the Mackie account of difference-making is similar to the kairetic criterion, and in particular, to the preliminary version of the kairetic criterion that I call the eliminative procedure. The main differences are, first, that the kairetic criterion's optimizing procedure represents a more sophisticated characterization of what makes a condition non-redundant, and second, that for Mackie, a set of conditions is sufficient for *e* if the conditions logically entail *e*, while on the kairetic account, they must causally entail *e*.

The most famous weakness of the unadorned INUS account—its apparent inability to deal with barometer/storm cases, such as Mackie's example of the Manchester factory hooters—is due to its putting no further constraint on the entailment. Mackie attempted to deal with the problem by requiring something a little like causal entailment (Mackie 1974, chap. 7). The INUS account is therefore in many ways a precursor of the kairetic account (Strevens 2004, in press b).