drical spacetime would yield the entropic behavior that we are familiar with. Consequently we cannot tell if it is more 'difficult' to engender our second law data in a closed world than in an open one. Therefore we have no reason to count these data as evidence against circular time.

These points carry over to Gödel's models. Just as in the case of the cylindrical spacetime, it is necessary that the initial state of a universe with closed timelike lines possesses a certain order. Moreover those special initial conditions that will *both* conform to a Gödelian spacetime *and* engender the entropic behavior we observe, form a small subset of all the possible initial conditions compatible with our entropic data. However, in the absence of any uncontroversial way of attributing a probability distribution to initial conditions, we cannot infer that the conditions needed to engender the second law data in a Gödel universe are unlikely. Thus we cannot conclude, in light of these observations, that there are probably no closed timelike curves. In addition, it is unclear that the second law data would be engendered by a smaller proportion of Gödelian initial conditions than non-Gödelian initial conditions. Therefore we cannot even claim that the second law data reduce the chances that we are in a Gödelian world.

Thus the argument against there being *any* time travel into the past is much weaker than the bilking argument against time travel into the local past. In the latter case specific coincidences would ensue, in whose improbability we can have a great deal of confidence. So we can conclude that such trips will not take place. However, this fact is explainable without having to posit a non-Gödelean spacetime. Therefore, as Gödel said, our own history is indeed out of bounds; but, for all we know, there do exist closed timelike lines, making a real possibility of trips to the spatially distant past.

8
Causation

1. *The problem of causal order*

The aim of science may be presented in a variety of ways: to discover laws of nature, to explain phenomena, to identify their causes, to say what would transpire in a range of hypothetical circumstances. None of these characterizations is especially controversial. Everyone will agree that causation, explanation, law of nature, and counterfactual dependence are methodologically vital and intimately related. Disputes arise, however—perhaps the central issues in philosophy of science—as soon as one tries to say precisely what these things are and precisely how they interact with one another.

This clutch of problems will be the general background theme of our next three chapters. More specifically, I will examine the respects in which causation, explanation, and counterfactual dependence are asymmetric in time. But the narrower project cannot be pursued without some attention to broader questions about the nature and affiliation of the concepts involved. Therefore I shall be proposing a rough picture of their internal structure and interrelationships—one that gives a primary role to the concept of explanation and characterizes causation, law, and counterfactual dependence in terms of explanation. My accounts of these notions will, I am afraid, be somewhat crude and not thoroughly defended. The point is to say just enough to sustain my explanations of their temporal properties.

Let me begin with the question of why effects rarely, if ever, precede their causes. This question is not the same as the problem of backward causation, discussed in chapter 6. There we took for granted that the typical direction of causation is toward the future, and asked whether there are any actual or possible deviations from this norm (a cause occurring later than its effect). The answer was no and yes—there aren't any actual cases as far as we know, but their existence is perfectly conceivable. The matter before us now is, rather, why the *predominant* direction of causation goes from past to future. This problem would remain even if there were occasional cases of backward causation.

The reason that the direction of causation is puzzling, and its explanation controversial, is this. On the one hand, causation appears to be a type of *determination*; on the other hand, all the types of determination with which we might plausibly identify causation seem to be time-symmetric. Consider, for example, the relation, "*C* is part of a condition that, given our laws of nature, requires the occurrence of *E*". This holds when *C* is later than *E* just as often as when *C* is earlier than *E*. And apparently the same thing goes for other prospective relations of determination—such as necessity, contiguity, or probabilistic connection—no matter how strong or weak they may be, and no matter how they are combined.

The natural—and, I think, correct—response to this situation, is to conclude that causation shouldn't be identified merely with some species of determination but with determination plus something else. A further ingredient must be added that will explain why causes tend to be earlier than their effects. But then the question arises as to what the extra condition should be, and this is the locus of much of the controversy surrounding the problem. Should we perhaps simply add the stipulation, "*C* is earlier than *E*"? This is the view I have called "conventional predetermination", since it treats time order as an a priori, analytic constituent of causation. Or should we, instead, add a different element to the definition of "causation"—one that happens, as a matter of contingent fact rather than as a matter of meaning, to constrain the time order of cause and effect. This strategy is realized in various alternative views, which I'll call "substantive" accounts, whereby the concept of causation does not in itself engender a time asymmetry but does so only together with other general facts about the world—facts that, had they been different, could have resulted in an opposite direction of causation, facts such as the fixed and settled character of past events (Mackie 1974), the prevalence of decay processes (Reichenbach 1956; Dummett 1964; Papineau 1985), the alleged temporal asymmetry of counterfactual dependence (Lewis 1979b), the orientation of action (Gasking 1955; von Wright 1971; Healey 1983), and the perception of time order (Mellor 1981).

I am going to argue that each of these two general points of view—conventional predetermination and substantive accounts—contain important aspects of the truth and that the right strategy is to combine them in a certain way. To see, in a preliminary fashion, what I have in mind, it is necessary to attend to the distinction between *explaining a fact* and *explaining why we should believe it.* In particular, we must separate the questions, "Why is it the case that causes typically precede their effects?" and "On what basis do we maintain that causes typically precede their effects?" These problems are not always clearly enough distinguished. (Perhaps this is because their answers have been thought to be the same.) According to the conventional predetermination picture, the *fact* is explained trivially by noting that time order is a constituent of causation, and the belief is explained as a priori. According to each of the substantive accounts, the fact is explained by whatever contingent features of the world are involved in generating the direction of causation from whatever definition of "causation" is assumed. And our belief in the future orientation of causation is derived a posteriori from our knowledge of those contingent features of the world. My account will extract pieces from each of these approaches. I will argue that the theory of conventional predetermination has roughly the right answer to why causes typically precede their effects, but that the substantive views give roughly the right explanation of why we ought to believe this. That is to say, I think the direction of causation is explained by noting that time order is a 'constituent' of the causal relation (in a sense to be described later), but this fact is a posteriori.

In presenting this view of the matter, I am going to distinguish between our *concept* of causation (which I take to be the cluster of all of our 'important' beliefs about causation) and our *theory* of causation (a subset of those beliefs—specifically, the subset that allows an especially simply specification of when the relation obtains). I will begin by proposing a certain neo-Humean theory of causation—a theory according to which it is constitutive of the relation of causation that causes typically precede their effects. Then I will turn to the explanation of why we believe this theory, particularly its temporal component. I will show that it can be derived, a posteriori and in several independent ways, from other elements of our concept of causation. Thus responsibility for our committment to the future orientation of causation is spread among many of the ideas that philosophers have thought were the source of it; there is no need to choose among them. In this way my explanation differs from that given by each individual substantive account.

Much of what I shall say about causation is obtained by beginning with the conventional predetermination view and modifying it in two directions. First, we should not think that every analysis of "*X* causes *Y*" in terms of time must be just the conjunction of some determining relation and the further condition '*X* is earlier than *Y*'. More sophisticated theories are possible in which the *earlier than* relation is involved in a complex way, but which still entail that causation is predominantly future oriented. If such a theory of causation is correct, it will explain the normal direction of causation and yet not be embarassed by the odd case of simultaneous causation or epiphenomena.

Second, any claim to the effect that something is 'true by convention', 'true by definition', or 'a priori' will be subject to Quine's (1951) powerful arguments against the viability of those notions. Therefore we must look for an account that can accommodate the ways in which all our commitments have some a posteriori character. To that end we should appreciate that a predetermination account can perfectly well be construed as an *a posteriori* theory rather than an *a priori* conceptual analysis. True, most proposals have been intended as accounts of the latter sort—as claims about the meaning of the word "cause". However, it is possible instead to offer a predetermination account as an a posteriori analysis of causation—an analysis that is intended to be somewhat analogous in status, for example, to the chemical analysis of water—describing the underlying structure of a phenomenon whose familiar symptoms merely provide a moderately reliable guide to when the stuff is present. Similarly I will suppose that the predetermination theory of causation serves to identify, in a precise and authoritative way, the relation that is the subject of the numerous more or less accurate maxims that constitute our concept of causation.

2. *A neo-Humean theory of causation*

The most influential proponent of building time into causation was David Hume, who characterized a *cause* as

> . . . an object precedent and contiguous to another, and where all the objects resembling the former are placed in like relations of precedence and contiguity to those objects that resemble the latter. (*Treatise*, Bk. 1, Part III, sec. XIV)

In other words

> C causes E *if and only if* every event like C immediately precedes an event like E

This theory has been improved in the light of some valid objections. Thus (1) it is plain that cause and effect need not be contiguous. Separated events, such as an explosion and the lighting of a fuse, may perfectly well be causally connected when there exists a causal chain from one to the other. (2) Events of different types might invariably follow one another by mere coincidence, and not in virtue of any causal relation between them: for instance, Mr. Smith's withdrawing $666 from his account, and his being struck by lightning as he leaves the bank. This sort of purely accidental correlation need not be especially unlikely if the events are so narrowly characterized as to be very rarely exemplified. (3) Striking a match sometimes causes it to burst into flames and sometimes doesn't, so being a cause of something does not entail constant conjunction with it. (4) Any two events are like one another in some ways, and not in others. Therefore the notion of similarity (which is needed to define 'like causes' and 'like effects') is vacuous in the absence of any indication of the relevant respects of similarity.

These problems have inspired various neo-Humean ('regularity') theories (e.g., Hempel 1965, p. 349; Popper 1972, p. 91). Here is a fairly promising version:

> A direct cause of some effect is an essential part of an antecedent condition whose intrinsic description entails, via basic laws of nature, that the effect will occur. And causation in general involves a chain of direct causation. That is to say, C causes E if and only if there are some events, $e1$, $e2$, . . . , eN, such that C directly causes $e1$, $e1$ directly causes $e2$, . . . , and eN directly causes E.

Briefly then, one can accommodate objection 1, as Hume actually did, by distinguishing direct and indirect causation and dropping the contiguity requirement (for the latter, at least); objection 2 is set aside by invoking the concept of law, and subsequently addressing (as we shall in chapter 10) the question of which regularities are to count as laws; objection 3 is dealt with by allowing that a mere part of the total cause may qualify as a cause; and as for objection 4, the significant respects of similarity are implicitly specified as intrinsic characteristics relevant to basic laws of nature. We require that the descriptions be *intrinsic* (in the sense discussed in chapter 3) in order to preclude the use of event characterizations that contain information about other events. Otherwise, a cleverly contrived description of C (e.g., a sneeze, described as "the sneeze that occurred twenty seconds before Caesar was stabbed") could, together with laws of nature, entail the occurrence of a causally unrelated event E (Caesar's death), merely by incorporating information about some event (the stabbing) that does cause E.

The preceding discussion is, of course only a sketch. Each move and countermove deserves a fuller treatment than I can give here without straying too far from the question at issue. I mention them only because my account of causal direction is set within the context of a neo-Humean regularity theory, and so it is important to indicate that such an account of causation is a live option. True, there are difficulties—but no compelling reason, I think, to suppose that they can't be overcome.

The account requires that the laws be explanatorily *basic* in order to distinguish direct from indirect cases of nomological determination, and thereby to be in a position to deal with certain standard objections, as we shall see. Our picture of direct determination is roughly as follows. When one type of event, C, nomologically determines another type E, then there is often a third event D such that the determination of E by C is explained by the determination of D by C and the determination of E by D. Such explanatory relations are most clear when the law $L1$ involved in going from C to D is different from the law $L2$ needed to get from D to E. In that case it is obvious that the nomological relation between C and E—namely, $L1$ & $L2$—is explained by the two more basic laws, $L1$ and $L2$. However, even if only a single law is involved in the process, there is still a sense in which the law relating C and E is explained by the law taking us from C to D, and from D to E. For example, suppose there is a law stating that whenever an event of type x occurs at spacetime region r, then an event of type $O(x)$ must occur at spacetime region $F(r)$. Now suppose that C is an event of type x occurring at $r1$, D an event of type $O(x)$ occurring at $F(r1)$, and E an event of type $O(O(x))$ occurring at $F(F(r1))$. Then, although there is some sense of "involvement" in which the same law is involved in going from C to D as in going from C to E, there is another sense in which different laws are involved. For, unlike the case of C and D, the nomological relation of which C and E are *instances*, and in virtue of which C determines E, is the following *derived* law: if an event of type x occurs at r, then an event of type $O(O(x))$ must occur at $F(F(r))$. In general, as we continue to interpolate events between determining events, we encounter relations that are more and more basic. In the limit we find relations of determination by means of the most basic laws, and these we identify with the elementary 'links' in a determination chain. Consider, for example, a law that specifies some aspect of the state of a system as a function of the state's initial condition and the elapsed time

$$St(t) = f(St(0), t)$$

If time were discrete, then we could reduce this law to more basic terms, so that it specified the state at one time in terms of the state at the preceding time

$$St(t+1) = g(St(t))$$

However, since time is continuous, the most direct cases of determination are those in which the state at a given time determines the state at an infinitesimally different time, and the most basic laws are the differential equations that describe this form of determination:

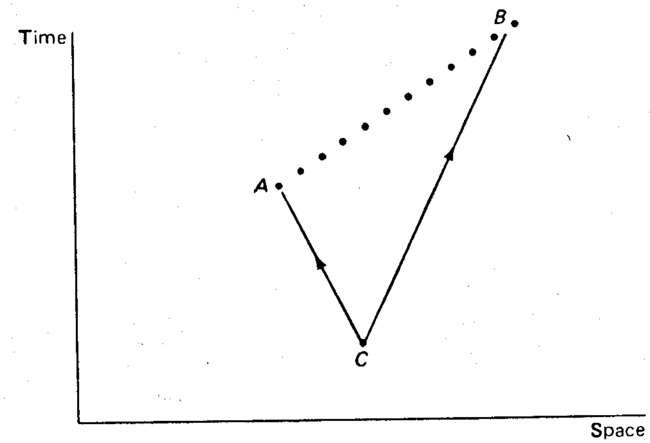

Figure 31

$$St(t+dt) = g(St(t)) \quad \text{or} \quad \frac{dSt}{dt} = h$$

I hope that these sketchy remarks help to make tolerably clear the idea of *direct* determination via *basic* law. Later on, some further light will be shed on 'laws' and what makes them 'explanatorily basic'.

One virtue of defining causation in terms of chains of direct determination is that we are able to avoid confounding cause and effect with mere epiphenomena. The latter are events that are nomologically related, not because one causes the other but through being common effects of the same cause. In such cases an event may determine a later event and yet not cause it. For example, a flash of lightning does not cause the clap of thunder that follows, and a change in barometer reading does not cause the subsequent storm. Schematically, cases in which A and B are epiphenomenal effects of C (as in figure 31) are distinguishable by the fact that the determination of B by A is explained by the existence of relations of determination between some earlier event C, and A, and between C and B. Thus the determination chain between A and B passes through C. So the only way for A to cause B would be for A to cause C and C to cause B. Therefore, given the absence of backward causation, it must be that C causes A and B.

A second advantage of defining causation in terms of chains of direct determination is that it helps us to solve a problem presented by simultaneous causation. The trouble is that, as it stands, the predetermination theory requires that a direct cause be earlier than its effects, and it therefore precludes simultaneous causation. But this is

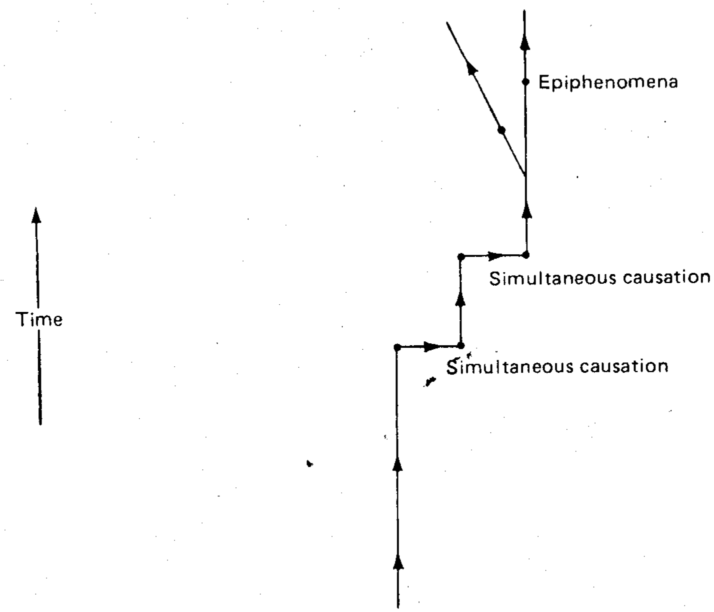

Figure 32

objectionable; for we are perfectly familiar with cases of causes being simultaneous with their effects—pushing or pulling, for example, where the application of a force to an object causes a simultaneous acceleration of that object.

In order to solve this problem, the role of time in the theory must be changed slightly. We should not impose the time-order requirement on each basic link in a causal chain. Instead, we should consider *entire* chains of determination (stretching between the distant past and the distant future) and then impose on the elements of any such chain a causal interpretation subject to the following pair of global contraints: (1) require that causes precede their non-simultaneous effects, and (2) maximize causal continuity (so that causal priority in one part of the chain may be 'smoothly' extended to adjacent parts). In other words, we begin with a chain of direct nomological determination strung out through time and perhaps containing occasional simultaneous links, as in figure 32. We now wish to associate 'arrows', representing the causal relation, with the basic parts of the chain, and in order to do so, we deploy constraints 1 and 2. We find, first, that the continuity condition tells us to avoid causal interpretations in which adjacent links have arrows that point in opposite directions. So that condition severely constrains the number of possible overall

causal interpretations of the chain. And this number is then narrowed down to a single possibility when we impose the requirement that causes precede their non-simultaneous effects. Thus time order is invoked to determine the usual direction of causation, in such a way that cases of simultaneous are not automatically precluded.

Backward causation is of course still ruled out. But note that this feature of the theory is by no means unfortunate, given that, as far as we know, there is no such thing. One might well have objected to our theory, were it put forward as an a priori definition. For no doubt backward causation is conceivable. But since the theory purports merely to describe a posteriori the actual nature of causation, it is not threatened by the fact that under different evidential conditions we would be inclined to take a different view of the matter.

I hope to have made it plausible in the last few pages that certain notorious objections to the Humean approach may be met. Thus I am resisting the view, propounded, for example, by Lewis (1973b), that problems arising from epiphenomena and backward causation torpedo the Humean approach and encourage us to look for an alternative account of causation in terms of counterfactual dependence. Not only is there little reason to give up on regularity analyses, but when we examine the counterfactual alternative (in chapter 10), we shall encounter difficulties that are, I think, even more substantial than those we find in the present approach to causation.

One final point—for we shouldn't leave the theory of causation without saying at least something about the question of *probabilistic* causation. Is it really necessary for causes to *determine* their effects? Here I have in mind not the issue of whether what we call "laws" are anything more than extremely widespread regularities but rather whether causes lead *invariably* to their effects. Hume's opinion—that they do, that given the same causes, the same effects will occur—is embodied in our practice of distinguishing causal from statistical laws and in formulating "determinism" as the thesis that every event has a cause. His idea is reflected in our analysis by the requirement that the effect be *deducible* from laws and initial conditions. On the other hand, there are occasions on which we recognize causes that neither determine nor play any role in determining what we regard as their effects (Anscombe 1971). For example, an overexposure to radiation might well be said to cause an illness, even when it is believed that this result was to some degree purely random.

My own view of this matter is that there is a genuine ambiguity at the bottom of it: the word "cause" has two meanings, one involving the notion of invariability and the other not. However, I won't presuppose that this is correct. Rather, I shall try to explicate a weaker conception of causation, and I shall assume that the stronger one may be obtained by adding a requirement of determination. What will matter, for our purposes, is that the weaker explication provide *necessary* conditions for causation that explain the direction of causation. Whether or not those conditions are also sufficient for *the*, or merely for *a*, concept of causation will not be important for the issue at hand.

One way of modifying our analysis to obtain an appropriately weaker conception is to let the circumstances of the alleged cause include not only surrounding initial conditions but also facts about the upshot of random processes. In other words, if the initial conditions, basic probabilistic laws of nature, and specific facts about how, given those initial conditions, the indeterministic factors will turn out together entail the occurrence of the event, then we shall say that those conditions directly caused the event. Thus suppose (rather implausibly) there is a basic statistical law that implies that people exposed to a certain radiation overdose stand a 20 percent chance of contracting leukemia. And suppose Smith gets both the overdose and leukemia. We may conclude (in the absence of any overdetermining factors) that the radiation caused the disease.

I have characterized the structure of causation by starting from a form of nomological connection and adding requirements of causal continuity and time order. Given this theory of what causation is, the explanation of why causes typically precede their effects is simply that this feature is *constitutive* of the causal relation—which is tantamount to saying that the fact has *no* explanation. For, in general, the theory that describes ~~the~~ nature of a type of entity may be deployed to explain further properties of the type, but the theory itself is not subject to explanation. Thus one does not ask why water contains oxygen or why planets orbit starts. Similarly no distinct fact is the explanatory basis for the direction of causation. To this extent my 'explanation' parallels that of the conventional predetermination model. Therefore it is equally susceptible to a certain standard objection, which I would like to discuss next.

3. Manipulation

A common charge is that if the direction of causation is to explain, as it should, the blatant irrationality of acting for the sake of the past, then the direction of causation cannot result simply from the presence of a temporal constituent in the causal relation. More specifically, Dummett (1954), Flew (1954), and Mackie (1974) have felt that if, for example, a cause were nothing more than an *earlier* necessary condition, then it would be arbitrary and wrong to require of a rational act

that it be intended to *cause* a desirable event. For there would be no reason why it wouldn't be just as good for one's act to 'retro-cause' a desirable event—where a 'retro-cause' is simply a *later* necessary condition. Therefore the predetermination model allegedly leaves us unable to explain why we do not do things now to ensure the occurrence of past events.

My response to this argument is to deny the presupposition that in order to be rational, an act must be intended to cause, rather than to 'retro-cause', a desirable event. It might be thought that in the absence of some such restriction the door would be wide open to wholesale action for the sake of the past. But that would be a mistake. For, even without the restriction, a past-oriented decision could be motivated only in very exceptional circumstances. To justify this claim, let me sketch a line of thought that will be fully developed in chapter 9. A past-oriented act may be rational only if we believe that it is *needed* to 'retro-cause' some earlier desirable event—that is, only if we are not already aware of some state that will retro-cause the event. And this condition will almost never be satisfied. For the chains of determination between our actions and past events are mediated by our motivational states of belief and desire. Therefore there is usually no potential retro-effect of a prospective action that we cannot *already* know will, or will not, be retro-caused by attention to our beliefs and desires. The only exceptions will occur in very contrived, hypothetical situations, such as Newcomb's decision context. And in such cases it is far from clear that past-oriented action actually is irrational. Thus the principle, "Act only for the sake of what might be caused," isn't needed to explain why we don't, in practice, act for the sake of the past. And in that case we can agree with Dummett, Flew, and Mackie that this principle does not square well with the predetermination-chain theory of causation. But, unlike them, we can regard the tension as grounds for rejecting, not the theory of causation, but the causal principle of rational choice.

The idea that there is an intimate connection between causal priority and rational motivation has been elevated to the status of an analysis of causation in the work of Gasking (1955) and von Wright (1971). Their 'manipulability' proposal is that the content of a causal generalization "C causes E" is something like "C could be used as a means for producing E", or "C may be directly controlled as a method of indirectly controlling E". But what can this possibly mean? If the analysis stops here, we are left with notions, like 'means' and 'method', that are more obscure than the one we were trying to explain. But, as Judith Thomson (1977) has pointed out, if we continue in the natural way, we would have to admit that what the criterion says is "There could be an action W such that W causes C and C causes E"—a construal whose reference to causation evidently disqualifies it as an analysis of that idea. The basic problem seems to be that the manipulability analysis is taking us in the wrong direction. It is more natural to explain the notions 'means' and 'method' in terms of causation rather than the other way round.

This problem is avoided in a formulation discussed by Richard Healey (1983, p. 92):

> . . . of two causally related events X and Y, if a person can produce an event of type Y without producing an event of type X, but not vice versa, then X causes Y.

However, this is tantamount to the view that any sufficient but unnecessary condition is a cause. It suffices to consider cases like the dryness and lighting of a match to see that it will not quite do as it stands. For in some circumstances a match can be dried without lighting it, but not vice versa; yet surely the lighting does not cause the dryness as Healey's criterion would seem to imply.

A better way of understanding the role of manipulability, in my opinion, is to regard it as providing one of the cluster of crude, fallible reference-fixing principles that help us to zero in on the causal relation. Thus we might roughly identify causation as "that relation believed to hold between rational choices and the desired events for the sake of which those choices are made". This maxim picks out a *future*-oriented relation. For, as we have just seen, it is typically irrational (for reasons independent of causation) to act for the sake of the past.

In my view, the manipulability maxim is ~~not~~ strictly accurate. There are occasions when we *should* act for the sake of events that we do *not* hope to cause, as we shall see when we examine Newcomb's problem in chapter 11. However, this departure from the truth shows merely that this maxim cannot be thought to capture the essential nature of causation. But it is not thereby disqualified from inclusion in the cluther of beliefs that makes up our concept of causation. It may perfectly well be a defeasible element of the concept, helping us to pick out a relation of causation whose underlying structure will then be described, in quite different terms, by the predetermination-chain account.

Thus I am supposing that the relationship between the concept and theory of causation is something like the relationship between our concept and theory of other parts of the world. For example, *water* is not constituted and identified by means of a set of analytically necessary and sufficient observable properties. Rather, it is recognized by

fallible criteria, such as 'colorless, tasteless liquid' and 'constituent of rain'. Its nature as H_2O is discovered afterward, and is then used to correct the original principles of identification. Similarly in the case of causation we should allow that its structure, as given by the predetermination-chain theory, may conflict with the symptoms by which we recognize instances of causation. These symptoms need be neither essential, universal, nor permanent. In particular, it could be that causation is identified in a loose way by maxims including (M)—"Causation is the relation believed to hold between rational choices and the desired events for the sake of which those choices are made"—that help to point us toward causation but do not describe its nature. The relation so ostended can then turn out to have a structure in which predominant time order is built in. And having arrived at this view of causation, we might then acknowledge cases (e.g., Newcomb's decision context) in which our initial maxims of identification are violated.

4. *Why believe that cause precedes effect?*

So far I have focused on the *metaphysical* question of why causes typically precede their effects. And I have answered it by suggesting a theory of causation in which time order is a constituent of the causal relation. It now remains to address the *epistemological* question of why we are right to *hold* that causes typically precede their effects. This is worth our attention for two reasons. First, it is important to combat the tendency to conflate the questions. (One often comes across attempts to explain the direction of causation in terms of the nature of our concepts.) Second, the answer to the epistemological question is what distinguishes the view presented here from the conventional predetermination account.

We have just encountered one way of reaching the conclusion that our belief in the direction of causation in a posteriori. Regard the manipulability maxim, (M), as an analytic principle partly defining "causation", and combine it with the belief that rational choices tend to *precede* the events for the sake of which the choices are made. These principles, the second of which is a posteriori, entail that causation tends to be future oriented.

Let us consider a more popular route to the same conclusion. Suppose it often happened that light contracted in inwardly moving spherical wave fronts to points in space—the time reverse of what actually occurs. And imagine that apples formed out of the earth and leapt up into the branches of trees. How could these processes be explained? Could mere coincidence account for the remarkable coor-

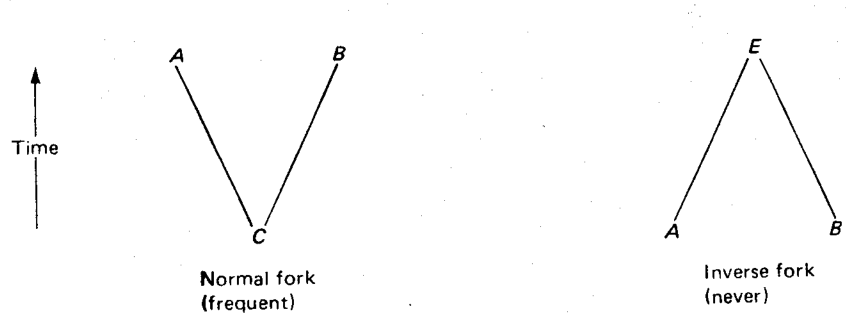

Figure 33

dination of separated wave fronts converging from infinity, or for the fact that apples are pushed from the ground by apparently random molecular vibrations in just the right directions and at just the right speeds to attach themselves to the conveniently receptive foliage? A tempting alternative would be to explain these things teleologically, by means of the hypothesis that the light waves are *destined* to be absorbed at a certain point and the apples *fated* to end their days in trees. Thus we might be a little inclined to postulate a form of backward causation.

This temptation comes from applying the familiar maxim, "Correlated events are causally connected," to a hypothetical pattern of events, known as an "inverse fork" (see chapter 4)—a pattern involving separated event types that are strongly correlated with one another and with a later event but not associated with any central, common antecedent. In the diagram on the right of figure 33, *A* and *B* represent parts of an inwardly moving wave front and *E* represents their ultimate absorption. In view of the symmetry of the situation, and the fact that the maxim of causal correlation is extremely well-entrenched, we are tempted to say that *E* causes *A* and *B*, rather than admit an uncaused correlation between *A* and *B*.

Thus one might invoke the maxim of causal correlation to explain our inclination to postulate backward causation. And in a similar way one might use it to explain why we believe in the predominance of forward causation. (Something like this idea has been endorsed by Reichenbach 1956, Dummett 1964, Mackie 1974, and Papineau 1985) Begin with the thesis that the direction of causation, in general, is fixed by the direction of causation in the special context of forks; then note that the maxim of causal correlation entails that in a fork the central event causes the separated correlated events; and, finally, bring in the vital contingent fact that there are many cases of so-called "normal forks" in which separated, simultaneous correlated events

are associated with an *earlier* central event but not with any characteristic *later* event, but the time reverse of this pattern—an inverse fork—is never exemplified. In other words, the direction of causation is analytically defined as the direction that would provide correlations with causal explanation. This is the direction in which forks spread out. And that, as a matter of contingent fact, happens to be the future.

Thus the collection of alternative 'substantive' accounts offers us a range of different a posteriori answers to the question of why we believe in the future orientation of causation:

1. "Causation" is defined—not in terms of time order—but, in part, by the principle that correlated events are causally connected, and this, given the fact that there are no inverse forks, determines that causation is future oriented.
2. "Causation" is defined through its association with our experience of deliberation and control. More specifically, we define causation as that general relation between events that is exemplified when an event is deliberately brought about by free choice. But, because of the difficulty of identifying the antecedents of decision, our voluntary actions are performed only for the sake of future events. Hence the direction of causation.
3. "Causation" is defined, in part, by the idea that a cause is in some sense 'ontologically more basic' than its effects. But, because of the knowledge asymmetry—very roughly speaking, the past is knowable and the future is not—we tend to think that the past has 'more reality' than the future (note our discussion of the tree model in chapter 2). And this leads to the idea that the past is causally prior to the future.

And perhaps there are more. But it is doubtful whether there is any clear sense to the thesis that one, and only one, of these answers can be correct. Here, I am endorsing Quine's idea that we ought to abandon the positivistic picture in which our beliefs about some phenomenon are divided into two disjoint groups, containing, on the one hand, those principles that merely define our terms and are therefore irrefutable and, on the other hand, those that express substantive, falsifiable, empirical claims. Instead, we should recognize that none of our beliefs is entirely sacrosanct, although some play a more pervasive, permanent, and prominent theoretical role than others. As our knowledge evolves, difficulties emerge in the current cluster of beliefs; some have to be given up. But where the axe falls is not constrained, as in the old picture, by some body of antecedent, once-and-for-all stipulations. Rather, the decision about how to proceed is shaped by the relative attractiveness of the various global theoretical positions that each option would result in. Applying this idea, we must renounce the view that some one of the preceding principles is part of the 'analytic definition' of "causation". Rather, we can suppose that the concept of 'causation' involves all of the principles on an equal epistemological footing.

Most scientific concepts are like this. Consider, for example, the classical idea of 'straight line'. One of its application criteria was 'satisfying the axioms of Euclidian geometry'; another, 'being the possible path of a light ray'; another, 'being the possible path of a freely moving particle'; and another, 'being the locus of a stretched string'. Within classical physics these are extensionally equivalent—all supposedly applying to the same entities. But with the collapse of that theory, the question arises as to which of these nonequivalent criteria pick out straight lines. In fact, what happened was that the geometrical criterion was abandoned and the others were preserved. Note, however, that this decision was motivated by considerations of theoretical economy within the new physics. There was nothing in our earlier, classical use of the straight line concept that showed that it would be right to make that decision. Therefore, in explaining why we believed that light travels in straight lines, it would have been wrong to cite convention and also wrong to pin responsibility on just one of the principles from which this belief may be derived. The right answer is to show how the belief is embedded in a large cluster of mutually supporting commitments.

I am suggesting that our attitude toward the direction of causation is similarly explained by its integrated position within a cluster of beliefs, including:

i. Causes typically precede their effects.
ii. Correlated events are causally connected.
iii. Choices are made for the sake of what they might cause.
iv. The causes of a present event are knowable, but its effects are not.

It is a contingent fact, partly a consequence of the scarcity of inverse forks and partly stemming from our experience of deliberation, that these elements are roughly consistent with superficial observation and seem able to coexist with one another. But, strictly speaking, they are not perfectly harmonious. We do have real knowledge of the future, we acknowledge that scientific advance could turn up occasional cases in which correlated events are not connected by prior causes, and, we come to see that in Newcomb-like decision contexts acts may reasonably be performed for the sake of earlier events. In

response to such facts the cluster of principles requires modification. Overall theoretical economy is promoted by recognizing that ii, iii, and iv are merely superficial approximations to the truth and by incorporating i within a characterization of causation along the lines of our predetermination-chain theory. However, although relegated to the category of "approximate truths", the maxims of correlation, deliberation, and knowledge remain part of our concept and play a significant role in bolstering the belief that causation is future oriented.