

CAUSATION

Chapter 14

Causation and the Direction of Time

14.1 Preliminary Considerations in Support of a Causal Analysis of Temporal Concepts

What reasons are there for thinking that at least those tenseless temporal relations that suffice to fix the order of events, and the direction of time, can be analyzed in a broadly causal fashion? There are, I think, at least four considerations. The first is that there appear to be striking structural similarities between the relations of temporal priority and causal priority. Both, it would seem, are necessarily irreflexive: an event cannot be earlier than itself, nor can it be causally prior to itself. Both are asymmetric: if A is earlier than B, then B cannot be earlier than A; if A is causally prior to B, then B cannot be causally prior to A. Both are transitive: if A is earlier than B, and B is earlier than C, then A is earlier than C; if A is causally prior to B, and B to C, then A is causally prior to C. Finally, one of the two directions associated with each of these relations has a special significance: it is *the* direction of time, or *the* direction of causation.

It is true that some philosophers have questioned whether temporal priority does have the properties just mentioned. Adolf Grünbaum,¹ for example, has contended that it is possible for time to be circular, and the argument that he has offered, if sound, would show that neither temporal priority nor causal priority need be asymmetric or irreflexive. Grünbaum's argument, however, rests on the principle of the identity of indiscernibles, against which there appear to be sound objections. In addition, there would seem to be a fairly firm, and quite widely shared intuition, that time cannot be circular. And, finally, any claim that causal priority need not be asymmetric or irreflexive is exposed both to the argument against the possibility of oppositely directed causal processes that I have set out elsewhere,² as well as to other arguments against causal loops, set out by Mellor and others.

¹See, for example, Adolf Grünbaum's discussion on pp. 614-19 of his article, 'Carnap's Views on the Foundations of Geometry', in Paul A. Schilpp (ed.), *The Philosophy of Rudolf Carnap* (La Salle, Illinois: Open Court, 1963), 599-684.

²*Time, Tense, and Causation*, (Oxford: Oxford University Press, 1997), chapter 4.

If causal priority and temporal priority do share these formal similarities, how can that fact be explained? A causal theory of time that postulates a merely contingent reduction of temporal relations to causal ones will not suffice, since the similarities would not then be a matter of logical necessity. What is needed, rather, is some sort of analytical connection between temporal and causal concepts.

That connection could run, of course, in different directions. Perhaps the formal similarities arise because causal concepts involve temporal ones, rather than because temporal concepts are analyzable in causal terms? Or perhaps they arise not because causal concepts presuppose temporal concepts, nor vice versa, but because both causal and temporal concepts are analyzable in terms of more fundamental concepts? In the absence of some concrete suggestion, I think that the latter possibility can be set aside. The former suggestion, however, certainly needs to be taken seriously, since many accounts of causation, such as Hume's, certainly involve reference to the relation of temporal priority. It seems to me, however, that a correct account of causation will not involve any temporal concepts.³ If this is right, then the most likely way for temporal and causal concepts to be analytically related is if the relevant temporal concepts presuppose causal ones. This, in turn, would seem to require either an analysis of temporal concepts in causal terms, or, at least, an analysis in which causal concepts play a crucial role.

A second, and related consideration is this. When a relation has certain formal properties, such as transitivity and asymmetry, one would like to be able to show why the relation has those formal properties. If one can analyze temporal concepts in terms of causal ones, one will *ipso facto* be providing at least a partial explanation of the formal properties of temporal relations. And if, as I have argued elsewhere,⁴ one can then go on to show why it is that causal relations have the formal properties that they do, the result will be a complete explanation of the formal properties of temporal relations.

A third consideration is that some statements that involve both causal and temporal concepts are widely regarded as expressing necessary truths - such as, 'A cause is always earlier than its effect'. If this is a necessary truth, what account of its necessity can be given? The most natural answer, it would seem, is that it is analytically true, and that this in turn is so because either causal concepts, when analyzed, involve temporal concepts, or vice

³Michael Tooley, *Causation :A Realist Approach* (Oxford: Oxford University Press, 1987), ch. 7, sec. 1, and ch. 8.

⁴*Ibid.*, ch. 8, sec. 4 and 5, and appendix. A different, and more general explanation of the formal properties of causation is provided, in effect, by the argument from causation set out earlier

versa. So, once again, provided that the former alternative can be rejected, there is a reason for concluding that the relevant temporal concepts can be analyzed in causal terms.

The fourth and final consideration arises when one asks what account is to be given of the direction of time. Many answers have been proposed, of course, most of them reductionist accounts, according to which the direction of time is given by certain patterns exhibited by events in time - such as the direction in which entropy increases, or the direction of the propagation of order in non-entropic, irreversible processes, or the temporal direction of the expansion of the universe. But there are strong reasons for thinking that the most that these answers can provide is a contingent identification of the direction of time with a direction defined in some other way, and not an analysis of the concept of the direction of time, since, given any reductionist account, one can describe possible worlds - including those I appealed to earlier in arguing against reductionist approaches to causation - where the account would fail to give the right answer. Thus, in the case of very simple worlds, for example, any account that defines the direction of time in terms of some pattern exhibited by events will entail, incorrectly, that time has no direction in such worlds, while, in the case of inverted worlds, such reductionist accounts will generate the wrong answer concerning the direction of time.

If such reductionist accounts are untenable, what alternatives are left? One possibility is a realist view, according to which the direction of time is an ultimate fact, not susceptible of any further explanation. But the problem with realism in this area is that it seems to provide no answer to the question of how one can have justified beliefs concerning the direction of time, since it gives one no reason for holding that the sorts of things that we normally take as relevant evidence - such as information about the direction of the propagation of order in non-entropic, irreversible processes - are indeed evidence. So this sort of account must, I believe, be rejected on epistemological grounds.

If a realist approach is unacceptable, and if the same is true for reductionist accounts that define the direction of time in terms of some pattern exhibited by events, it is not easy to see what possibility remains other than that of analyzing the direction of time in terms of the direction of causation, and then adopting a realist account of causation. But this alternative seems perfectly satisfactory, since, provided that one can have justified beliefs concerning the direction of causation, one will be able to have justified beliefs about the direction of time, while avoiding the objections which tell against reductionist accounts that identify the direction of time with a direction defined in terms of some pattern in the world.

14.2 An Absolute, or a Relational Account?

One important issue that needs to be addressed before setting out a causal theory of time concerns the choice between an absolute, or realist, or substantival view of space and time (or space-time), and a relational, or reductionist view. According to the former, talk about space and time (or space-time) cannot be reduced to talk about the spatial and temporal relations between physical objects and events. The totality of space-time points is itself a substance - that is, an entity that is logically capable of independent existence. Therefore, either that totality is something over and above physical objects and events and their spatiotemporal relations, or alternatively, as some absolute theories maintain, physical objects and events are themselves simply portions of space, or of space-time, possessing certain properties.

On the relational view, by contrast, talk about spatial and temporal relations is primary, and talk about space and time can be replaced, without loss, by reference to spatiotemporal relations between physical objects and events. Space and time, therefore, are nothing over and above physical objects and events, and their spatial and temporal interrelations.

Causal theories of time have almost invariably been relational in form. The reason, I think, is probably this. On the one hand, to accept an absolute view of space and time is to hold that space and time might exist even if there were no physical objects or events, while, on the other, to accept a causal theory of time is to hold that at least all tenseless temporal relations are analyzable in terms of causal relations. This means that, if one accepts both views, one has to hold that spatiotemporal regions can stand in causal relations to one another, since otherwise one will not be able to make any sense of the idea of a possible world in which space and time exist, but where there are no physical objects or events.

It is certainly true that there are conceptions of space and time that do involve the idea that spatiotemporal regions enter into at least some causal relations. According to the General Theory of Relativity, the curvature of space-time is determined by the matter present, and it, in turn, determines how bodies will move. But, if one considers instead the Newtonian picture, or an ordinary person's view of space and time, the situation is quite different. Spatiotemporal regions are thought of as not interacting at any time, either with physical objects, or with other spatiotemporal regions. And it is natural to conclude from this that spatiotemporal regions do not themselves enter into any causal relations. Accordingly, if one is concerned with the problem of making sense of our ordinary notions of space and time, it might seem that one cannot accept both a substantival view of space and time, and a causal theory of time.

But is this not a decisive objection to the idea that one can offer a causal analysis of temporal concepts? For, can it not be argued that we do believe that empty space and time are a logical possibility, that space and

time could exist even if there were no matter or energy at all, so that our ordinary conception of space and time, like the Newtonian, is substantival, and thus that, if a causal theory of time, and an absolute view of space and time cannot be combined, our ordinary temporal concepts cannot be analyzed in causal terms?

How can an advocate of a causal theory of time respond to this difficulty? The only satisfactory reply, I believe, involves showing that a substantival view of space and time is not in fact incompatible with a causal theory of time. But how can that be done? The answer seems clear: one must show that it is possible for spatiotemporal regions to stand in causal relations to one another, for then the logical possibility of a world devoid of matter poses no objection to a causal theory of time.

But is it really possible for there to be causal relations between spatiotemporal regions? One place to start, in thinking about this question, is by asking whether there are any conceptual constraints upon when it is logically possible for two contingent, non-simultaneous states of affairs to be causally related. The answer, I suggest, is that Hume was right on this matter: anything can cause anything. But, if this is right, if the concept of causation places no restrictions at all upon what types of contingent, non-simultaneous states of affairs can be causally related, then, given that the existence of a spatiotemporal region is a contingent state of affairs, it follows that it is logically possible for spatiotemporal regions to enter into causal relations.

Secondly, it is crucial here that one not equate the very general notion of causal *relationships* with the more specific idea of causal *interactions*. This is, I think, a real danger, since, when one talks about causal relations, what often springs to mind are causal interactions. Then, since it is not part of our ordinary concept of substantival space that parts of it interact with other things, the idea that there can be causal relations between spatiotemporal regions naturally seems unacceptable. But the situation changes, I suggest, when one keeps in mind the fact that relations of causal dependence are also present in cases where there is no interaction, or change, taking place. Consider, in particular, the relation of identity over time. If that notion is analyzable, causation will surely play a crucial role. But the causal dependence of later temporal parts of an enduring entity upon earlier temporal parts is clearly not a matter of causal interaction. And once one focuses upon this sort of situation, the idea that spatiotemporal regions can be causally related ceases, I suggest, to be problematic, since, if one accepts a substantival view of space and time, space itself is an enduring entity, and thus the idea that later temporal parts of space are causally dependent upon earlier ones is no more strange than the idea that later temporal parts of an electron are causally dependent upon earlier ones.

In short, if one focuses upon causal relations of the sort involved in the persistence of enduring objects, it becomes clear that spatiotemporal regions can stand in causal relations, and thus that there is no a priori objection to combining a causal theory of time with an absolute view of space and time. But are there any reasons for preferring such a formulation of a causal theory of time? At least three, it seems to me. First, if there are good reasons for accepting a realist, or substantival view of space-time, rather than a relational conception, one needs an account of temporal relations that will apply not merely to events, but to spatiotemporal regions as well. There appear to be, however, good reasons for accepting a realist conception of space-time. One involves Newton's famous thought experiments of the rotating bucket containing water, and of the two globes connected by a cord.⁵ Consider, for example, the latter. Newton's argument was that, if the two globes were rotating about the center of mass of the system, the cord joining them would be under tension. Yet the globes would not be in motion relative to one another. Moreover, the situation would not be changed by the supposition that no other objects existed beyond the two globes and the cord. So the tension could not be caused by the motion of the globes relative to one another, or relative to other material things. It could be caused only by the absolute rotation of the system - that is, its rotation relative to absolute space.⁶

Newton himself thought that this argument, in establishing the need for the idea of absolute rotation, and thus of acceleration relative to space itself, also showed that one had to accept the ideas of absolute velocity, and absolute location. The inference here is certainly a very natural one. However, as present-day philosophers such as John Earman have pointed out, it is mistaken: one can set out mathematical models of space-time that, while allowing for absolute acceleration, do not involve absolute velocity or absolute location.⁷ But, while Newton's argument does not establish the conclusion that he thought it did, it does provide a good reason for holding that a relational account of space-time is unsatisfactory.⁸

The other reason why one must, I believe, adopt a realist view of space-time involves the need to provide an account of certain empirical possibilities. In particular, consider the fact that there are locations where there *could*, at a given time, be a physical object, even though, as a matter of

⁵Sir Isaac Newton, *Mathematical Principles of Natural Philosophy* (Berkeley and Los Angeles: University of California Press, 1934), 10-12.

⁶Ibid. 12.

⁷John Earman, 'Who's Afraid of Absolute Space?', *Australasian Journal of Philosophy*, 48/ 3 (1970), 287-319. See pp. 296-7.

⁸Compare Earman, *ibid.*, p. 312.

fact, this is not the case. The possibility involved here is not that of a mere logical possibility. Our everyday experience of the motion of objects, for example, makes it reasonable to believe that there are locations where there are no physical objects, but where the existence of such objects is *empirically* possible. But then if, as I believe, such modal facts cannot be taken as ontologically primitive, one is confronted with the question of the categorical basis of these empirical possibilities. Accounts are certainly possible, of course, that do not involve a substantival view of space. It might, for example, be the presence of some physical field in an otherwise empty location which is the ground of the empirical possibility of there being an object at that location. Or perhaps the empirical possibility is somehow grounded in the categorical properties of physical objects, events, or fields that exist at other locations than the one in question. However, I think it is fair to say that both of these hypotheses are extremely speculative. The question, therefore, is whether there is any plausible alternative to the view that the categorical properties that are the basis of the empirical possibilities in question are properties of the locations themselves, and thus in no way dependent upon the properties of physical objects, or events, or fields - there or elsewhere. If not, then the rejection of a relational account in favor of a substantival one is necessary if there is to be any categorical basis for the empirical possibilities in question.

A second reason for embedding a causal theory of time in a substantival theory of space-time is connected with a serious problem confronting all causal theories of time - a problem posed by the issue of whether temporal relations are to be analyzed in terms of actual causal connections alone, or whether reference must also be made to the causal connections there would have been if things had been different. The difficulty here is that neither answer seems satisfactory. On the one hand, if one frames the analysis only in terms of actual causal connections between events, there may very well be possible cases where those relations will not suffice to fix, logically, the temporal relations between the events in question. On the other hand, appealing to hypothetical causal connections gives rise to the issue of the truth-makers for the relevant counterfactuals. This question would pose no problem, of course, if one could refer to spatiotemporal relations between events. But in a context of setting out a causal theory of time, such reference threatens to generate a circularity in the analysis, and, as a consequence, the question of the truth-makers for the relevant counterfactuals may become very difficult indeed.

A causal theory of time that is combined with a relational view of space and time lacks the resources, I believe, to handle this problem. But, if one adopts, instead, an absolute view of space and time, and holds, as seems reasonable, that every complete temporal slice of space is causally connected with every other, then those actual causal relations between temporal slices of the world, together with spatial relations within temporal slices, will suffice

for the assignment of a temporal location to every event. There will therefore be no need to appeal to possible causal connections, and the serious problem of providing truth-makers for such hypothetical connections will be completely avoided.

The final consideration involves a number of familiar objections to causal theories of time, among them the following. First, it has been argued that causal theories of time cannot make sense either of the possibility of space-time points where there are no events, or of the more dramatic possibility of times at which nothing exists. Secondly, it has also been objected that causal theories of time cannot allow for the possibility of events in space-time that are uncaused, and that have no effects. Thirdly, there is the objection that causal theories of time are incompatible with a possibility that is allowed by the general theory of relativity - namely, that of totally empty spatiotemporal worlds.⁹ Directed against relational causal theories, these are, I believe, very damaging objections. But, by contrast, and as we shall see later, they pose no problem for substantival versions of the causal theory of time.

14.3 Modal versus Non-Modal Analyses

Some causal theories of time involve the modal notion of two events being causal *connectible*, while others are formulated instead in terms of the non-modal notion of two events being causally *connected*. Are there reasons for preferring one of these approaches over the other?

If one wanted to set out either a causal theory of spatiotemporal relations in general, or a theory of temporal relations that referred to nothing other than causal relations, then it seems clear that one would have to go with causal connectibility, rather than causal connectedness, for surely there are pairs of events where one is earlier than the other, but which are not causally connected. Faced with such events, there would seem to be only two possible ways to formulate a causal theory: either one must refer not only to causal relations, but to spatial relations as well - as in the account set out below - or one must refer to possible causal relations, as well as to actual ones. The first of these, however, is not an option if one's goal is to set out either a purely causal theory, or a theory covering absolutely all spatiotemporal relations.

Let us begin by considering, then, the idea of an account that is formulated in terms of causal connectibility, rather than in terms of actual causal connections. The basic line of thought involved in such an approach to temporal relations might be put as follows. If two events are causally related,

⁹These objections are set out in J. J. C. Smart, 'Causal Theories of Time', in Eugene Freeman and Wilfrid Sellars (ed.), *Basic Issues in the Philosophy of Time* (LaSalle, Illinois: Open Court, 1971), 61-77.

then one is earlier than the other. But what if they are not causally related? It will not do, in that case, to say that they are simultaneous, for it may be that, although they are not causally related in the world as it is, they would have been if the world had been slightly different, and surely one does not want to allow that, had the world been different in that way, the two events would have stood in a different temporal relation. Accordingly, in a case where two events are not causally related, one has to ask whether they *could* have been so related. If they could have been, then one is earlier than the other, whereas, if it would have been impossible for them to have been causally related, then neither is earlier than the other.

The idea of shifting from actual causal connections to causal connectibility, or accessibility, is certainly a very natural one. A crucial question needs to be answered, however, before a connectibility account can be accepted - namely, what is it *in the world* that makes it the case that two events are, or are not, causally connectible?

Why is this question a pressing one? The answer turns upon a certain view concerning empirical possibilities, to the effect that, first, empirical possibilities are to be cashed out in terms of the truth of relevant subjunctive conditionals, and, secondly, that those subjunctive conditionals require truth-makers that consist of *categorical* facts, possibly together with laws of nature. If this view of empirical possibilities is right, so that it will not do to treat empirical possibilities as simply ultimate facts, not capable of further explanation, the question is what sorts of categorical facts constitute the basis of the causal connectibility of two events. One answer, of course, is that the categorical facts are a matter of spatiotemporal relations between the two events. So, for example, if one considers a relativistic case where *E* and *F* are two events that are not causally connected, but where *F* lies within the forward light cone of *E*, then the two events are causally connectible in virtue of that geometrical fact. But, if one's goal is to show either that all spatiotemporal relations are reducible to causal relations, or that temporal relations are reducible to causal relations alone, it obviously will not do to hold that the categorical facts that underlie causal connectibility involve spatiotemporal relations between events. The question then is: what alternative is there? What other type of categorical state of affairs can serve as the relevant truth-maker? I do not believe that there is any. But, if this is right, then a connectibility account cannot provide either a satisfactory analysis of spatiotemporal relations in general, or an analysis that reduces temporal relations to causal ones.

Does this mean that no connectibility account can be sound? Not quite. If one's goal is simply to show that temporal relations are reducible to causal relations *plus spatial relations*, one will not be involved in any circularity if one holds that the categorical facts that ground the relevant empirical possibilities involve spatial relations between events. But, if this is the sort of causal theory that one has in mind, is there any advantage in

formulating things in terms of causal connectibility? Might it not be better to formulate the theory in terms of actual causal connections, together with spatial relations, thereby making it clear from the beginning exactly what the reduction base for temporal relations is?

To sum up. Causal theories of temporal or spatiotemporal relations that are formulated in terms of the modal idea of causal connectibility may seem very natural and appealing, given that actual causal connections are not sufficient for defining the relation of temporal priority. But those who have put forward connectibility theories have failed to address a crucial issue - that of the categorical facts that ground the relevant empirical possibilities - and, once this issue is raised, it becomes clear that any connectibility theory that seeks to reduce either temporal or spatiotemporal relations to causal relations alone, either actual or possible, cannot succeed.

14.4 Qualitative Temporal Relations: Simultaneity and Temporal Priority

14.4.1 The Analysis of Simultaneity and Temporal Priority

A satisfactory account of tenseless temporal relations needs to cover both qualitative and quantitative relations. Here, however, I shall focus only upon the qualitative relations of simultaneity and temporal priority, and, for reasons that emerged in the preceding section, the analysis will be in terms of actual causal relations plus spatial relations.

According to our ordinary, non-relativistic conception of time, two relations suffice to order events in the universe: simultaneity and temporal priority. The former is, by definition, reflexive, symmetric, and transitive, while the latter is transitive, but asymmetric and irreflexive. In addition, those two relations are mutually exclusive, and, together with the inverse of temporal priority, exhaustive. So, given any two instantaneous events,¹⁰ A and B , it cannot be the case both that A and B are simultaneous, and that A is earlier than B , while, on the other hand, it must be the case that either A and B are simultaneous, or A is earlier than B , or B is earlier than A . It follows from these assumptions that the temporal ordering of events is a total ordering: absolutely all events in the universe are part of a single, linear order.

I now want to consider what happens when a certain very weak assumption concerning the connection between causation and time is conjoined with our ordinary conception of time. First, however, notice that one consequence of that conception, as characterized above, is this:

¹⁰Throughout the following discussion, events are restricted to *instantaneous* states of affairs.

- (P) If *A* is earlier than *B*, and *B* is simultaneous with *C*, then *A* is earlier than *C*.

The reason is that, in view of the exhaustiveness assumption, either (a) *C* is earlier than *A*, or (b) *A* and *C* are simultaneous, or (c) *A* is earlier than *C*. But, (a) is excluded, since it, together with the fact that *A* is earlier than *B*, would entail that *C* is earlier than *B*, and that, in turn, in view of the fact that simultaneity and temporal priority are mutually exclusive, would entail that *B* and *C* are not simultaneous. And, (b) is also excluded, since it, together with the fact that *B* and *C* are simultaneous, would entail that *A* and *B* are simultaneous, and the latter is precluded by the fact that *A* is earlier than *B*. So, if *A* is earlier than *B*, and *B* is simultaneous with *C*, then *A* must also be earlier than *C*.

A precisely parallel argument would establish that the following is also entailed by the above characterization of our ordinary conception of time:

- (Q) If *A* is simultaneous with *B*, and *B* is earlier than *C*, then *A* is earlier than *C*.

Suppose now that one introduces the following assumption:

- (R) If *A* causes *B*, then *A* is earlier than *B*.

This is an assumption that would be accepted in the vast majority of causal theories of time - with the exception of some more holistic theories which, in order to allow for the purported possibility of backward causation, identify the direction of time with something like the dominant direction of causal processes. But, if the argument of Chapter 4 is correct, backward causation is not logically possible, and, accordingly, *R* should lie at the heart of any causal theory of time.

R specifies one condition under which one event is earlier than another. Clearly, however, there must be other conditions that are also sufficient, since *A* can be earlier than *B* even if they are not causally connected. It is at this point, as was noted earlier, that the idea of shifting from talk of causal connections to talk of causal connectibility naturally arises. But, if that move is, as I argued, unsatisfactory, how should the case of causally unrelated events be handled? The answer emerges very quickly, I think, if one begins by noticing that acceptance of *R*, in the context of our ordinary conception of time, commits one to further propositions concerning conditions that are sufficient to ensure that one event is earlier than another. In the first place, *R*, when conjoined with *P* and *Q*, entails, respectively:

- (S) If *A* causes *B*, and *B* is simultaneous with *C*, then *A* is earlier than *C*;
 (T) If *A* is simultaneous with *B*, and *B* causes *C*, then *A* is earlier than *C*.

So we now have two other conditions that suffice to ensure that one event is earlier than another. But, in the second place, these two conditions, in conjunction with the fact that temporal priority is a transitive relation, entail another, much more encompassing condition:

- (U) If $\{A_1, A_2, \dots, A_i, \dots, A_{n-1}, A_n\}$ is a set of n instantaneous events such that, for every $i < n$, either A_i causes A_{i+1} , or A_i is simultaneous with A_{i+1} , and if, in addition, there is some $i < n$ such that A_i causes A_{i+1} , then A_1 is earlier than A_n .

Principle U , entailing, as it does, principles R , S , and T , and more as well, is a very comprehensive principle relating causation to temporal priority, and the fact that it follows from the conjunction of our ordinary conception of time with the very modest claim involved in R shows how strongly any causal theory of time is constrained by assumption R .

Principle U , however, like principles S and T , suffers from a certain obvious defect, if one's goal is to formulate a causal theory of time, since its specification of the conditions under which one event is earlier than another is not done in terms of causal and other non-temporal notions alone: the description involves the concept of simultaneity. What is needed, accordingly, is a principle that is comparable to U in power, but which does not involve, in the characterization of the conditions under which one event is earlier than another, any temporal notions.

Given an account of simultaneity, U could be transformed in the way desired. But what account can be offered of simultaneity? One natural answer is to analyze simultaneity in terms of causal unconnectibility: two events are simultaneous if and only if they are not causally connectible. But what possibility is there if causal connectibility cannot be employed? The only possibility, I think, is an account that connects simultaneity with spatial relations.

Does that mean that the concept of simultaneity is to be analyzed in terms of the idea of spatial relations? Perhaps. But, for the purposes of transforming U , it is not really necessary to advance a claim about the analysis of simultaneity. A weaker proposition will suffice, concerning only a sufficient condition for the simultaneity of two events, namely:

- (V) If events A and B are spatially related, instantaneous events then they are simultaneous.

But is V acceptable? Events that are not simultaneous can, after all, certainly happen, say, five miles apart. The answer is that when one speaks of spatial relations between things or events that do not exist at the same time, one must always bring in, implicitly or explicitly, the idea of being in the same place at different times, and so the question arises as to what account is to be given of that notion. In Chapter 11 I shall argue that

sameness of place is to be analyzed in terms of the presence of appropriate causal relations connecting spatiotemporal regions. If this is right, then, when one speaks of spatial relations between things or events that exist at different times, the relation being attributed, rather than being a *pure* spatial relation, involves a combination of pure spatial relations plus causal relations. Principle *V*, accordingly, must be viewed as being formulated in terms of pure spatial relations.

Given principle *V*, it is possible to derive a principle which, like *U*, specifies a very wide range of conditions under which one event is earlier than another, but which, unlike *U*, characterizes those conditions *in non-temporal terms*. For the conjunction of *U* and *V* entails:

(W) If $\{A_1, A_2, \dots, A_i, \dots, A_{n-1}, A_n\}$ is a set of n instantaneous events such that, for every $i < n$, either A_i causes A_{i+1} , or A_i is spatially related to A_{i+1} , and if, in addition, there is some $i < n$ such that A_i causes A_{i+1} , then A_1 is earlier than A_n .

To recap briefly. I have argued that two quite modest principles - namely, *R* and *V* - in conjunction with our ordinary conception of non-relativistic time, entail principle *W*. The first of those principles asserts that, if one event causes another, then it is also earlier - a proposition that, unless backwards causation is possible, would certainly appear to be true. The other principle asserts that spatially related, instantaneous events must be simultaneous - a claim that also seems very plausible. Accordingly, any account of our ordinary temporal notions would appear to be subject to the very strong constraint that is embodied in principle *W*.

W formulates a very comprehensive sufficient condition for one event's being temporally prior to another. But could the condition in question also be a necessary condition? In that case, the converse of *W* would also be true:

(W*) If A_1 is earlier than A_n , then there is a set of instantaneous events $\{A_1, A_2, \dots, A_i, \dots, A_{n-1}, A_n\}$ such that, for every $i < n$, either A_i causes A_{i+1} , or A_i is spatially related to A_{i+1} , and, in addition, there is some $i < n$ such that A_i causes A_{i+1} .

Does *W** express a necessary truth? If not, then I think that it is not easy to see how an analysis of temporal priority can be formulated in terms of actual causal connections, rather than in terms of causal connectibility. But then, given the difficulty that confronts the latter sort of account, it seems that a causal analysis of the earlier-than relation may very well stand or fall with the tenability of the claim that *W** is necessarily true.

Let us assume, then, that both *W* and *W** express necessary truths. This allows us to formulate the following definition of temporal priority:

A is earlier than *B*

means the same as

For some number n , there is a set of n instantaneous events $\{A_1, A_2, \dots, A_i, \dots, A_{n-1}, A_n\}$ such that, first, A is identical with A_1 , and B is identical with A_n ; secondly, for every $i < n$, either A_i causes A_{i+1} , or A_i is spatially related to A_{i+1} ; and, thirdly, there is some $i < n$ such that A_i causes A_{i+1} .

Similarly, the simplest way of extending principle V into an analysis of simultaneity is by assuming that the condition for simultaneity specified by it is necessary as well as sufficient, so that one can offer the following definition:

A is simultaneous with B

means the same as

A is spatially related to B .

14.4.2 Some Consequences of this Account

To determine whether these analyses are satisfactory, we shall need to consider - as I shall in a later section - how they fare in the face of objections. Two preliminary matters, however, can profitably be addressed at this point: first, whether simultaneity and temporal priority, thus defined, have the appropriate formal properties; secondly, whether propositions which involve both temporal and causal concepts, and which seem to express necessary truths, turn out to do so given the above accounts.

First, then, let us consider the formal properties of simultaneity and temporal priority. The case of simultaneity is straightforward. First, it is surely a necessary truth that the inverse of any spatial relation is also a spatial relation, so, if A stands in some spatial relation to B , then B must stand in some spatial relation to A . The relation of simultaneity, defined as above, is therefore symmetric. Secondly, it is surely also a necessary truth that, if A is spatially related to B , and B to C , then A is spatially related to C . So the transitivity of simultaneity is ensured.

What about temporal priority? First of all, there is no problem about transitivity. If A is earlier than B , and B than C , then there must be two chains of the appropriate sort - one from A to B , and the other from B to C - and then the combination of those two chains will make it the case that A is earlier than C .

Next, what about the asymmetry of temporal priority? At this point, a distinction drawn in Chapter 3 is crucial - namely, that between cases of backward causation involving at least the possibility of causal loops, and cases of causal processes running in opposite directions, but not involving the possibility of causal loops - for what that distinction shows is that, although one cannot establish the asymmetry of the relation of temporal priority, as

defined above, unless one can show that causal loops are logically impossible, the latter by itself will not suffice. One must be able to show that oppositely directed causal processes are also impossible. The argument that was offered in Chapter 4 in support of the latter claim is therefore crucial.

Secondly, what about propositions - such as the proposition that a cause is always earlier than its effect, or the proposition that an effect never precedes its cause - which involve both causal concepts and temporal ones, and which it seems plausible to view as expressing necessary truths? Do they turn out to do so, given the above analyses? The answer is that they do. Thus, the proposition that a cause is always earlier than its effect is an immediate consequence of the analysis of temporal priority. In the case of the proposition that an effect never precedes its cause, one also needs to appeal to the proposition that oppositely directed causal processes are logically impossible. Given that proposition, however, the result follows very quickly, since, if it were the case both that C causes E , and that E is earlier than C , it would follow from the analysis of temporal priority that there were oppositely directed causal processes. So an effect can never precede its cause.

In short, the situation seems promising, since we have seen that the qualitative relations of temporal priority and simultaneity can be analyzed in terms of actual causal relations, together with spatial relations, in a way that does not, at least initially, seem to be counterintuitive, that generates the appropriate formal properties, and that seems to entail necessary truths that involve both causal and temporal concepts.

14.5 Objections to Causal Theories of Time

Causal theories of time are exposed to a number of important objections, most of which fall into two groups. First, many philosophers have argued that, even if the analytical equivalences advanced by causal theories do obtain, those analytical truths do not provide satisfactory *analyses* of temporal concepts. Secondly, potential counterexamples have been offered to the claimed analytical connections between temporal concepts and causal concepts.

Most of this section will be devoted to addressing these two types of objection. In my response to objections of the second sort, the reader will notice that I make extensive use of the fact that the causal theory of time that I have advanced involves both an absolute view of space and time, and the idea that spatiotemporal regions can be causally related. It might be argued, however, that this enables me to avoid some standard objections to causal theories of time only at the cost of new difficulties. I shall therefore conclude my discussion in this section by addressing one argument of that sort which may seem especially damaging.

14.5.1 The Analytical Equivalences Do Not Provide Analyses

A number of reasons have been advanced for holding that, even if the definitions proposed by causal theorists are analytically true, they do not provide satisfactory definitions. Two of the more important are that the analyses are implicitly circular, and that involve intensional contexts.

14.5.1.1 Causal Priority Presupposes Temporal Priority

One of the main arguments offered for the view that causal theories do not provide satisfactory analyses of temporal concepts involves the claim that such theories are implicitly circular: when the proposed analyses are carefully scrutinized, it turns out that the concepts involved in the analyses presuppose temporal concepts.

One version of the circularity objection - mentioned by J. J. C. Smart in his paper 'Causal Theories of Time' - is directed against accounts that employ an asymmetrical causal relation: 'In the absence of a prior notion of earlier and later how can we distinguish cause from effect, especially in view of the time-symmetry of the laws of nature?'¹¹

Some advocates of a causal theory of time - such as Grünbaum, Mehlberg, and others - have tried in effect to finesse this objection by formulating the analysis in terms of a *symmetrical* causal relation, such as that of causal connectedness.¹² I do not believe, however, that this strategy is satisfactory. The crucial issue here is whether the most basic causal relations are symmetrical or asymmetrical. If a reductionist account of the direction of causation were sound, then it might well be possible to view asymmetrical causal relations as definable in terms of underlying symmetrical relations plus the relation that defines the direction of causation. But, since reductionist accounts of the direction of causation appear to be exposed to decisive objections, I believe that this approach is untenable. Analysis here must run in the opposite direction, with symmetrical causal relations, such as causal connectedness, being analyzed in terms of the fundamental, and asymmetrical, relation of causation. So the idea of formulating a causal theory of time in terms of a relation such as causal connectedness cannot really provide a satisfactory answer to the present charge of circularity.

¹¹J. J. C. Smart, 'Causal Theories of Time', in Eugene Freeman and Wilfrid Sellars (ed.), *Basic Issues in the Philosophy of Time* (LaSalle, Illinois: Open Court, 1971), 61-77. See p. 62.

¹²Adolf Grünbaum, *Philosophical Problems of Space and Time*, 2nd ed. (Dordrecht: D. Reidel, 1973), ch. 7, and Henryk Mehlberg, 'Essai sur la théorie causale du temps', *Studia Philosophica* 1 (1935), 119-258 and 2 (1937), 111-231.

It would seem, then, that only a direct response will do: one needs to argue that a sound account of causation, and, in particular, of the direction of causation, will not involve any appeal to temporal relations. But, as will be clear from the discussion in Chapter 4, I believe that such an account can be offered.

14.5.1.2 Modal Concepts and Categorical Facts

A second and, I believe, much more forceful type of circularity objection focuses upon modal concepts such as causal connectibility, and causal accessibility. Here the thrust is, first, that the relevant causal concepts are modal ones; secondly, that modal facts require a categorical basis; and, thirdly, that the relevant categorical basis for the causal facts in question will involve, in some way, spatiotemporal relations.

One version of this objection, for example, turns upon the idea that whether two events that are not in fact causally connected are causally connectible depends not only upon what laws there are, but also upon how the events are spatiotemporally related. But then one cannot understand what it is for two events to be causally connectible, unless one first has the concept of a spatiotemporal relation. Consequently, any analysis of spatiotemporal relations that involves notions such as causal connectibility must be circular.

A somewhat different, though closely related version of the present objection focuses upon the laws that enter into the states of affairs in virtue of which events are causally connectible. Thus Smart, after observing that the modal concept of causal connectibility presupposes reference to laws of nature, points out that 'if these laws of nature themselves presuppose the very structure of space-time which we are seeking to elucidate by means of the notion of causal connectibility we are clearly involved in a vicious circularity'.¹³

Both of these objections tell, I believe, against many formulations of a causal theory of time. But they have no force against the account set out above, since it is formulated in terms of actual causal connections - including connections between spatiotemporal regions - rather than in terms of the modal notion of causal connectibility.

14.5.1.3 Intensional Contexts

A third important consideration offered in support of the contention that causal theories of time cannot provide satisfactory analyses of temporal

¹³J. J. C. Smart, 'Causal Theories of Time', in Eugene Freeman and Wilfrid Sellars (ed.), *Basic Issues in the Philosophy of Time* (LaSalle, Illinois: Open Court, 1971), 61-77. See p. 62.

notions turns upon the claim that such theories involve modal, or at least, intensional, concepts. The thrust of this objection is that, since, for example, truth is not necessarily preserved by the substitution of co-extensive expressions in sentences involving intensional contexts, the truth conditions of such sentences are very far from perspicuous, and, as a result, any analysis couched in terms of such concepts cannot be acceptable. To analyze temporal concepts in terms of causal ones is, accordingly, to analyze the less obscure in terms of the more obscure.¹⁴

The claim that modal and intensional notions cannot be taken as analytically basic seems right to me, and so, for example, in the case of causal theories of time based upon the idea of causal connectibility, I would insist that it is crucial to show that that idea can be explained in purely extensional terms. Moreover, given that the concept of causal connectibility involves the idea of laws of nature, this in turn means that one would need to show that one can provide a satisfactory, purely extensional account of the truth conditions of statements expressing laws of nature.

But how does this objection bear upon a causal theory of time that is formulated in terms of actual causal connections, rather than in terms of causal connectibility? The answer is that some writers, such as Smart, have suggested that the idea of causal connectedness is itself an intensional notion,¹⁵ and, given that some causal statements do involve intensional contexts, that claim is not without some initial plausibility. But there is a perfectly satisfactory response to this version of the objection - namely, that one can formulate a purely extensional language to talk about properties and relations, and, if one uses that language to describe causally related states of affairs, the resulting causal statements will be completely extensional. Non-extensional contexts occur in causal statements only if one confines oneself to a nominalistic language - where there are no terms that refer to properties and relations.

14.5.2 Possible Counterexamples

Objections of the second sort involve attempts to show that it is logically possible for there to be temporally related events that do not stand in the relevant causal relations. These objections do refute, I believe, most causal theories of time, but not the version advanced above.

14.5.2.1 Empty Spatiotemporal Regions

The thrust of this objection is that there could be space-time points that are not occupied by any events, but that, given a causal theory of time, there

¹⁴Ibid. 63-6.

¹⁵Ibid. 63.

will be no way of assigning temporal location to such space-time points, since there will be no events at such points to stand in causal relations to anything.¹⁶

A closely related objection, mentioned to me by J. J. C. Smart, concerns the more dramatic possibility of times at which nothing at all exists. That this is indeed a possibility needs to be argued, of course. But I believe that this can be done by arguing that action at a temporal distance is logically possible. In particular, if one can establish that temporally gappy causal laws are logically possible, then it would seem that one can show that there could be a universe that, although it contained objects at most times, was sometimes completely devoid of objects, since by choosing an appropriate set of temporally gappy causal laws, plus initial conditions, one could describe a universe in which, on occasion, the temporal gaps in all of the laws would happen to coincide, with the result that absolutely nothing would exist at those times.¹⁷

Both of these possibilities pose a serious problem for relational versions of the causal theory of time. Neither has any force at all, however, against causal theories formulated in terms of causal relations between space-time points, rather than merely between events in space and time.

14.5.2.2 Events that Are Not Causally Connected to Other Events

A second objection, also due to J. J. C. Smart, focuses upon the possibility of 'points of space-time which are occupied by events which are neither effects nor causes of other events'.¹⁸ This possibility also constitutes a very serious problem for any relational version of the causal theory of time, since, while initially it might seem that such events could be handled by appealing to possible causal relations, the earlier discussion of causal connectibility makes it clear, I think, that an advocate of a relational view will not, in general, have resources that will suffice to supply categorical truth-makers for the relevant modal statements in the case of events that are, by hypothesis, causally completely isolated. But, once again, if one opts instead for a causal theory of time that is formulated in terms of causal relations between space-time points, no difficulty is presented by the possibility of causally isolated events.

¹⁶Compare *ibid.* 62.

¹⁷This argument is very closely related to one advanced by Sydney Shoemaker in 'Time without Change', *Journal of Philosophy*, 66 (1969), 363-81.

¹⁸*Ibid.* 63.

14.5.2.3 The Logical Possibility of Empty Space-Time

A third objection appeals to the extreme possibility of spatiotemporal worlds in which no events take place at any space-time point. If such totally empty spatiotemporal worlds are logically possible, then all relational versions of the causal theory of time are ruled out. But are such worlds really possible? In support of the claim that they are, some philosophers have appealed to the fact that there are solutions to the equations of General Relativity that describe worlds possessing a space-time structure, but devoid of all matter and energy.¹⁹ It might very well be argued, however, that such an appeal needs, at the very least, to be supplemented by a consideration of the arguments which advocates of a relational view of space-time have advanced for thinking that the idea of empty space-time is, in the final analysis, incoherent.

Is there is a good reason for thinking that a totally empty spatiotemporal world is not logically possible? I do not believe that there is, and I shall shortly attempt to provide support for this claim. For the moment, however, let us simply assume that a totally empty spatiotemporal world is logically possible. The present objection will then be a decisive argument against relational versions of the causal theory of time. But it poses no difficulty for theories that are formulated in terms of causal relations between space-time points, since such theories are perfectly compatible with the possibility of a world where space-time is totally empty.

14.5.2.4 Continuous Causal Processes and Quantum Mechanics

A fourth objection, also raised by J. J. C. Smart, turns upon the fact that causal theories of time need to postulate the existence of continuous classes of events in space-time. The problem is that, while it is true in classical physics that between any two genidentical events there will be a continuous class of genidentical events, the situation is quite different in the case of quantum physics, where it is hard even to make much sense of the assumption that there are such continuous classes of events. The absence of definite trajectories, for example, points to 'an uncertainty as to just what is supposed to be the set of events which would make up a genidentical causal chain'.²⁰

¹⁹Ibid. 67-8. Also Henryk Mehlberg, 'Relativity and the Atom', in Paul Feyerabend and Grover Maxwell (ed.) *Mind, Matter, and Method - Essays in Philosophy of Science in Honor of Herbert Feigl* (Minneapolis: University of Minnesota Press, 1966), 449-491. See especially pp. 484-6.

²⁰J. J. C. Smart, 'Causal Theories of Time', in Eugene Freeman and Wilfrid Sellars (ed.), *Basic Issues in the Philosophy of Time* (LaSalle, Illinois: Open Court, 1971), 61-77. See p. 67.

Can this objection be brought to bear upon the causal theory advanced above? To do so, one would have to show that quantum physics is compatible with a discrete space-time. But, as Smart emphasizes, the latter idea is very dubious:

Quantum mechanics uses continuous mathematics and continuous geometry, like any classical theory. Indeed it is not at all clear what a physical geometry based on a discrete space-time would look like, especially in view of the well known incommensurability of certain geometrical ratios and in view of Zeno's paradox of the stadium.²¹

It would seem, therefore, that the world as described by quantum mechanics does not constitute an objection to theories of time that invoke causal relations between space-time points themselves.

14.5.2.5 The Possibility of Eternal Recurrence

A fifth objection to causal theories of time centers upon the possibility of a universe involving eternal recurrence - that is, which consists of an unending sequence of qualitatively indistinguishable temporal segments. Why does this possibility pose a problem for causal theories of time? The reason is as follows. Suppose that such a world contains events S , T , and S^* such that, first, S causes T ; secondly S and S^* are corresponding events in distinct, but qualitatively indistinguishable temporal segments, so that S and S^* are indistinguishable with respect to all of their properties, both intrinsic and relational; and, thirdly, S^* is later than T , and does not cause T . The question, then, is how it can be the case that S causes T , whereas S^* does not. If one is not committed to a causal theory of time, this question presents no problem, for then what makes it the case that S causes T , while S^* does not, can be simply the fact that S is earlier than T , while S^* is later than T . But, if one holds that temporal relations are analyzable in terms of causal relations, this answer is unavailable, since it would render the analysis circular.

This objection is perfectly general, and applies to all causal theories of time, including the one advanced here. It is also a very serious objection, and rather desperate countermeasures have sometimes been taken. Consider, for example, Adolf Grünbaum's discussion. His response to the apparent possibility of 'a universe consisting of a platform and one particle constantly moving in a circular path without friction' is that it is 'inadmissible' to interpret this as a case where 'the same kind of set of events (circular motion) keeps on recurring eternally' - a contention he supports by appealing to the Leibnizian principle of the identity of indiscernibles: 'if two states of the world have precisely the same attributes, then we are not confronted by distinct states but merely by two different names for the same state at one

²¹Smart, 'Causal Theories of Time', 63.

time'.²² Grünbaum's answer, in short, is that the supposed possibility is not genuine.

I do not think that Grünbaum's response is the right one, for it seems to me that familiar objections to the Leibnizian principle of the identity of indiscernibles - such as that of the possibility of a world that consists of an infinite number of qualitatively indistinguishable objects, spaced an equal distance apart - are decisive. But what is the alternative? The answer is that one can reject the fundamental assumption concerning causation upon which the argument rests - namely, the following principle of Humean supervenience:

Whether or not E causes F is logically fixed by the non-causal properties of E and F , plus the non-causal relations between E and F , together with what causal laws exist.

Given this principle, the argument is clearly sound. First of all, S and S^* are, by definition, the same with respect to all of their properties, both intrinsic and relational. Secondly, though S is earlier than T , while S^* is later than T , if temporal priority is analyzed causally, there is no non-causal relation, R , such that S , but not S^* , stands in relation R to event T . Finally, we are dealing with a single possible world, so there is no difference with respect to what causal laws there are. Consequently, if Humean supervenience obtains, it follows that S can cause T if and only if S^* also causes T - contrary to the description given of the possible world in question.

This argument collapses if the above thesis of Humean supervenience is rejected. But, given how widely accepted Humean supervenience is, a rejection of it may well seem no less desperate than Grünbaum's appeal to the principle of the identity of indiscernibles. I believe, however, that there are decisive arguments against Humean supervenience, one of which - the argument from the possibility of indeterministic laws - was set out in Chapter 4.²³ The correct response to the present objection, accordingly, is simply that the thesis of Humean supervenience is false.

14.5.2.6 The Possibility of Global Causal Loops

A final objection appeals to the possibility of universes containing a global causal loop, with an event at one time giving rise to a sequence of

²²Adolf Grünbaum, p. 614 of 'Carnap's Views on the Foundations of Geometry', in Paul A. Schilpp (ed.), *The Philosophy of Rudolf Carnap* (La Salle, Illinois: Open Court, 1963), 599-684.

²³For a statement of the other arguments, see pp. 222-33 of 'Causation: Reductionism versus Realism', *Philosophy & Phenomenological Research*, Suppl. 50 (1990), 215-36.

causally linked events such that the original event is itself a later member of that sequence. But what reason can be offered for thinking that such worlds are really possible? Sometimes the appeal is an 'innocent until proven guilty one', and it is argued that the attempts that have been made to show that global causal loops are impossible are not in fact sound.²⁴ But appeals are also frequently made to the fact that such a spatiotemporal world involving a global causal loop 'can be given a description internally consistent and consistent with the field equations of general relativity'.²⁵

Why does the possibility of such a world present a difficulty for causal theories of time? The basic thrust of the argument is that the analytical connections between temporal concepts and causal ones postulated by a causal theory of time generate absurd consequences when applied to such a world. Thus it will follow, for example, that some events occur after themselves, or that some pairs of events are such that each is both before and after the other. One is forced, therefore, to reject the alleged analytical connections between temporal concepts and causal ones.²⁶

Even if one were to grant the possibility of such causally cyclical universes, I think that the force of the present objection would be far from clear, since, if one thought that such a world were possible, would not one also think that temporal priority was not an asymmetric relation, and thus that there was no absurdity in the conclusion that an event could be earlier than itself? But there is no need to pursue this issue. If, as I have argued, oppositely directed causal processes are impossible, then the world described, however consistent it may be with the field equations of General Relativity, is not a logically possible world.

14.6 Objections to a Realist View of Space and Time

In responding to many of the objections set out above, I appealed to a realist conception of space and time, and one in which spatiotemporal regions themselves stand in causal relations. This strategy has made it possible to avoid a number of difficulties, many of which would otherwise be decisive objections to a causal theory of time. One might well view this maneuver, however, as a case of buying into even more serious problems. In the present section, accordingly, I shall consider some objections to realist views of space and time in general, and to the specific version that I have advanced.

²⁴See, for example, Susan Weir, 'Closed Time and Causal Loops: A Defence Against Mellor', *Analysis*, 48 (1988), 203-9.

²⁵Lawrence Sklar, *Space, Time, and Spacetime* (Berkeley and Los Angeles: University of California Press, 1974), 303.

²⁶Compare Sklar, *ibid.*, 335.

14.6.1 Leibnizian-Style Arguments against Absolute Space

What objections can be mounted against a realist, or substantialist, view of space and time? In his book *Space, Time, and Spacetime*, Larry Sklar suggests that there are three main types of argument - all deriving from Leibniz's correspondence with Samuel Clarke: (1) epistemic or verificationist arguments; (2) arguments based upon some version of the principle of the identity of indiscernibles; and (3) arguments based upon the principle of sufficient reason.²⁷ Given the implausibility of the principle of sufficient reason, I shall ignore the third type of argument. The other two types, however, need to be considered.

Sklar formulates the verificationist argument as follows:

The meaningful assertion of the existence of some entity or feature of the world requires that the presence or absence of that entity or feature, or a change in that feature, have some observational consequences. To affirm the existence of features of the world with no detectable consequences is not to espouse some kind of meaningful skepticism, but rather to affirm the intelligibility of the unintelligible. Now suppose space itself were a substance. It would then make sense to ask what the position of the whole of the material world in space is, how fast the world as a whole is moving with respect to this substantialist space, etc. But we can only observationally determine the spatial relations of material objects relative to one another, the motions of material objects relative to one another, etc. There are no observations that could conceivably determine the position of the world as a whole in substantialist space, nor its velocity with respect to substantialist space, etc., assuming of course that, for example, in changing its position in substantialist space the internal spatial relations of material objects relative to one another remain constant. So now we see that: (a) Belief in substantialist space requires the intelligibility of assertions about the position of the world as a whole in substantialist space and its motion with respect to substantialist space; but (b) such assertions are clearly meaningless by the verificationist principle. Therefore, there can be no such thing as substantialist space.²⁸

There is not much to be said about this argument, since it stands or falls with the verificationist principle. But, given that there are very good reasons for thinking that the verifiability principle is unsound - including the fact that it implies that realism with regard to theoretical entities is an unintelligible position - I believe that the above argument can justifiably be set aside.

The final type of argument rests upon a version of the principle of the identity of indiscernibles. Consequently, this second type of argument might seem unpromising, given that standard formulations of the principle of the identity of indiscernibles are exposed to familiar, and apparently quite convincing, counterexamples. Sklar points out, however, that the usual formulations of the principle deal with the identity of objects, and that there

²⁷Ibid. 173-80.

²⁸Ibid. 173-4

is another version of the principle that deals, not with the identity of objects, but with the identity of possible worlds:

*P.3. Suppose we have possible worlds A and B such that they are the same with regard to every purely qualitative feature. Then A is the same possible world as B.*²⁹

Sklar suggests that this principle is much more promising than standard formulations of the principle of the identity of indiscernibles in that, first, it is not exposed to the usual counterexamples, and, secondly, it is possible to 'offer some reasons for believing that *P.3* might be logically true'.³⁰

Given this principle, the argument against a substantival view of space then runs as follows:

Consider the two possible worlds that consist of (1) the actual world and (2) the actual world displaced five feet to the north in substantival space. It seems as though there are two possible worlds being talked about here, but the worlds are identical with respect to every purely qualitative feature. So they are, by *P.3*, the same possible world. So substantival space does not exist.³¹

If principle *P.3* were correct, I think that this would be a sound argument. However, I believe that it is possible to construct a counterexample that shows that *P.3* is false. Consider a world that, throughout all past time, has contained nothing except an infinite number of motionless, equally spaced, qualitatively indistinguishable billiard balls, all in a straight line. Pick out one of the balls, and call it 'X'. Now imagine that only one change occurs in this world throughout all of time, namely, at some moment, every other billiard ball vanishes without a trace. This could happen in one of two ways, since X might or might not be among the billiard balls that disappear. Let us call these two possible worlds 'A' and 'B'. It is clear that A and B are distinct worlds, since one contains X throughout its history, while the other contains X only during its initial stages. Yet worlds A and B are the same with regard to all purely qualitative features. So principle *P.3* is false.

14.6.2 Causation, Absolute Space, and Humean Supervenience

The thrust of this final objection is that the idea of causal relations between spatiotemporal regions is incompatible with the thesis of Humean supervenience mentioned earlier. To see why, consider a completely empty, spatiotemporal world that contains three space-time points *P*, *Q*, and *R* such that *P* is earlier than *Q*, and simultaneous with *R*, and the existence of *Q* is caused by the existence of *P*, but not by the existence of *R*. If spatiotemporal

²⁹Ibid. 176.

³⁰Ibid. 177.

³¹Ibid. 178.

regions and points can be causally related, the above must surely be a logically possible world. But is it? The problem here is that, if the principle of Humean supervenience is true, then the existence of Q can be caused by that of P , but not by that of R , only if either P differs from R with respect to its non-causal properties, or else Q stands in some non-causal relation to P that it does not stand to R . But space-time points do not differ with regard to their intrinsic properties, and, given that we are considering an empty universe, neither can they differ with respect to their relational properties. So the only way that the existence of P can be a cause of the existence of Q , while the existence of R is not, is if there is some non-causal relation that obtains between P and Q , but not between R and Q . What could that relation be? Since P and R are simultaneous, it cannot be any temporal relation. It is hard to see what possibility there is, then, other than that P stands in a different spatial relation to Q than R does. But this involves the postulation of absolute spatial relations holding between things that exist at different times, and the claim that there can be such relations is widely held to be unintelligible - as is shown by the almost universal rejection of Newton's idea of location in absolute space. It seems, then, that any spatial relation that holds between P and Q , must also hold between R and Q . Consequently, P and R neither differ with respect to any non-causal properties, either intrinsic or relational, nor stand in any different non-causal relations to Q . The principle of Humean supervenience then implies that it is impossible for the existence of Q to be caused by the existence of P , while not being caused by the existence of R .

This argument does appear to me to establish an interesting conclusion - namely, that, if the principle of Humean supervenience is true, there cannot be causal relations between spatiotemporal regions or points. For although the argument, as it stands, does involve an assumption that I believe is incorrect - the assumption, namely, that one cannot make sense of Newton's idea of location in absolute space - the way in which, I would argue, one can make sense of this notion is in terms of causation: two space-time points are in the same location if and only if they are causally connected. Consequently, the above argument will not really be undercut by this point, since, although it implies that Q can stand in a different spatial relation to P than it does to R , that difference will not be a non-causal one.

In short, it seems that we have here a sound argument for a conditional conclusion, and, if one adds the additional premise that the principle of Humean supervenience is true, it will follow that there cannot be causal relations between spatiotemporal regions. But, as will be clear from earlier discussion, that additional premise is, I believe, untenable.

14.7 The Pervasiveness of the Direction of Time

I have argued that there are positive considerations that point in the direction of a causal theory of time, and that such a theory, when properly

formulated, does not succumb to standard objections. I shall now conclude my defense by mentioning a feature of causal theories of time that seems to be an important and attractive one, and which competing theories of time virtually never possess.

The point has to do with the basis of the direction of time, and it can be developed by considering a property that the direction of time has on a number of non-causal analyses - such as that of Popper, on the one hand, and those of Reichenbach and Grünbaum, on the other - but which it does not possess given a causal analysis. The property in question is the dependence of the direction of time upon initial or boundary conditions. Consider, for example, Grünbaum's description of his own approach to the problem of the direction of time:

We must first describe certain features of the physical world having the character of initial or boundary conditions within the framework of the theory of statistical mechanics. The sought-after basis of a statistical anisotropy of time will then emerge from principles of statistical mechanics relevant to these *de facto* conditions.³²

This dependence of the direction of time upon initial or boundary conditions has struck a number of philosophers as unsatisfactory. Thus Henryk Mehlberg, for example, commenting on the idea that the direction of time can be defined by reference to such things as the fact that the incidence of expanding optical spheres far exceeds that of contracting ones, says:

Once more, however, the decisive point seems to be that the asymmetry between the two types of light waves depends on factual initial conditions which prevail in a given momentary cross section of cosmic history or at the 'boundaries' of a finite or infinite universe rather than on nomological considerations concerning this history: any other ratio of the incidences of expanding and shrinking light waves would also be in keeping with the relevant laws of nature contained in Maxwell's theory of electromagnetic phenomena.³³

Mehlberg is not very explicit as to exactly why he thinks it is objectionable to explain the direction of time in such a way that it is dependent upon factual, initial conditions. His feeling seems to be, however, that, if time has a direction, it must be a pervasive feature of the whole physical universe, rather than something that would have been radically altered by a change in a momentary cross section of the world.

This feeling about the pervasiveness that the direction of time ought to possess leads Mehlberg, in turn, to conclude that time has no direction, that it

³²Grünbaum, p. 651 of 'Carnap's Views on the Foundations of Geometry', in Paul A. Schilpp (ed.), *The Philosophy of Rudolf Carnap* (La Salle, Illinois: Open Court, 1963), 599-684.

³³Henryk Mehlberg, 'Physical Laws and Time's Arrow', in Herbert Feigl and Grover Maxwell (ed.), *Current Issues in the Philosophy of Science* (New York: Holt, Rinehart and Winston, 1961), 105-38. See pp. 123-4.

is isotropic. His reason is, first, that he apparently believes that the only way in which the direction of time can be a pervasive feature of the world is if it is grounded in laws of nature, and, secondly, that it appears to be the case that the laws of nature are, as a matter of fact, invariant under time reversal.

Both Mehlberg's insistence upon the idea that the direction of time must possess a pervasive quality, and his view that this feature is not compatible with the idea that the direction of time is dependent upon initial or boundary conditions, seem plausible. By contrast, however, his contention that the direction of time can have the desired quality only if it is grounded in the existence of laws of nature that are not invariant with respect to temporal reversal seems to me to be mistaken, in view of the features that the direction of time will have, given the approach developed above. First, time's having a direction is perfectly compatible with all the laws of nature being invariant with respect to temporal reversal. Secondly, the direction of time is nevertheless not dependent upon initial or boundary conditions. Thus, in contrast to what is the case given the approaches of Popper, Reichenbach, Grünbaum, and others, the direction of time would not be changed, for example, by an alteration in the distribution and/or velocities of particles in some momentary cross-section of the physical universe. Nor is the direction of time a statistical property. If there is a direction to time, it is identical with the direction of causation in absolutely every causal process, including not only those involved in the continued existence of fundamental particles, but those involved in the continued existence of space-time itself. As a consequence, the direction of time is about as pervasive a fact as there could possibly be - involving, as it does, *every* causal relation between states of affairs.

It is possible, in short, to set out an account of the direction of time according to which it is not dependent upon initial or boundary conditions, without appealing, as Mehlberg believes one must, to laws of nature that are not invariant with respect to time reversal. Causal theories of time, including the one advanced here, do precisely this. They also imply that the direction of time is an extraordinarily pervasive fact about the world. This combination of features is, I suggest, a very strong consideration in support of a causal theory of time.

14.8 Summing Up

The causal analysis of temporal concepts advanced above has a number of desirable features. First, the analysis entails that temporal relations do possess the formal or structural properties that they are normally believed to have: the relation of simultaneity is transitive, reflexive, and symmetric, while the relation of temporal priority is transitive, irreflexive, and asymmetric.

Secondly, there are striking structural similarities between the relations of temporal priority and causal priority, and, in providing a derivation of the formal properties of temporal relations on the basis of corresponding properties possessed by causal relations, causal theories of time provide an explanation of those similarities.

Thirdly, there are propositions involving both causal concepts and temporal ones - such as the proposition that a cause is never later than its effect - which it seems plausible to view as being necessarily true. Causal theories of time provide a straightforward explanation of that necessity: the propositions in question are analytically true.

Finally, many non-causal theories of temporal direction entail that the direction of time depends upon initial or boundary conditions. This feature seems undesirable, for it seems very natural to view the direction of time as an extremely pervasive feature of the world, rather than as a feature that would disappear if the initial conditions were changed in certain ways. A causal theory of time, by contrast, ensures that the direction of time is such a pervasive fact, for, wherever there is causation, there is temporal direction. This pervasiveness is accentuated, moreover, if one adopts the view that I have urged here, according to which spatiotemporal regions themselves stand in causal relations to other spatiotemporal regions, since then the direction of time is part of the very fabric of space-time itself.