CAUSATION

Chapter 13

The Question of Backward Causation

How might one attempt to show that it is logically impossible for an effect to precede its cause? In this section I shall briefly describe six important arguments that have been offered. The first appeals to the idea that causation involves the bringing of effects into existence; the second, to a connection between causation and control; the third, to the possibility, given backward causation, of bringing about contradictions; the fourth, to the problem posed by the possibility of undercutting causal chains; the fifth, to the view that causes must render their effects more likely; and the sixth, to the idea that the existence of causal laws involves a 'transmission' of probabilities from causes to effects.

13.1 Causes as Bringing Effects into Existence

Arguments for the claim that an effect cannot precede its cause can be divided into those that either presuppose, or directly support, a dynamic view of time, and those that are neutral between dynamic and static views. One example of the former sort appeals to the ideas, first, that the concept of causation is that of the *bringing into existence* of one event by another; secondly, that the bringing into existence of an effect by its cause means that, on the one hand, as of the time of the cause's occurrence, the effect does not exist, whereas, on the other hand, as of the time of the effect's occurrence, the cause does exist; and, thirdly, that if one event is either earlier than, or simultaneous with, another event, then the former does exist as of the time of the latter.

The basic objection to this argument is that it does not seem plausible that what one might called the idea of 'tensed succession' is part of *the very analysi*s of the concept of causation, since many philosophers both hold that the world contains causal processes, and either accept a tenseless view of time, or else a tensed view, but one according to which past, present, and future events are all equally real.

It is possible, of course, that while the idea of tensed succession is not part of the very analysis of the concept of causation, one can establish that causation can only exist in a world where there is tensed succession. Indeed, I have attempted elsewhere to show that that is so.[1] But the argument is a complicated one, with premises that are by no means uncontroversial, and I shall not consider that argument here.

13.2 Causation and Potential Control

The preceding argument incorporates a dynamic conception of time in a completely explicit fashion. This second argument, by contrast, initially appears to be neutral with respect to tensed versus tenseless views of the nature of time. I shall contend, however, that the argument in question quickly collapses unless a tensed view of time is invoked.

The starting-point for this second argument is that there is an analytical connection between the concepts of causation and control: causes are necessarily such as could, in principle, be used to control the occurrences of events that they cause; they are, in principle, effective means of bringing events of the relevant sort into existence. Given that this is so, consider the hypothesis that backward causation is possible. In particular, consider the idea that an event of type C at location s_2 at time t_2 would be causally sufficient to bring about an event of type E at location s_1 at an earlier time t_1. Once t_1 is past, one could in principle know whether an event of type E had occurred at location s_1 at a time t_1,. But then it can be argued that regardless of whether an event of type E has or has not occurred, bringing about an event of type C at location s_2 at time t_2 cannot be an effective means of bringing it about that an event of type E occurred at location s_1 at time t_1. For if an event of type E has not occurred, then it will be impossible to bring it about that it has occurred by bringing about an event of type C, whereas, on the other hand, if an event of type E has occurred, then the later production of an event of type C is unnecessary, and therefore cannot be an effective means of bringing about an event of type E at the earlier time t_1. So, either way, the production of an event of type C at a later time t_2 cannot be an effective way of producing an event of type E at an earlier time t_1.[2]

---

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

[2]For formulations of this argument, see, for example, Antony Flew ,'Can an Effect Precede its Cause?', *Proceedings of the Aristotelian Society*, Suppl. 28 (1954), 45-62, especially pp. 57-8, and Max Black, 'Why Cannot an Effect Precede its Cause?', *Analysis*, 16 (1955-6), 49-58, especially p. 58.

Both of the crucial claims just made can, however, be challenged. First, consider the claim that, if no event of type *E* occurs at time t_1, then bringing about an event of type *C* at a later time t_2 cannot be an effective means of bringing about an event of type *E* at the earlier time. It is true that, if no event of type *E* occurs at location s_1 at time t_1, then it is nomologically impossible for an event of type *C* to occur at location s_2 at the later time t_2. But how does this entail that production of an event of type *C* is not an effective means of producing an event of type *E*? For might it not be true that, if there *were* to be an event of type *C* at location s_2 at time t_2, there *would* have been an event of type *E* at location s_1 at time t_1? And, if that counterfactual were true, would not the production of events of type *C* at later times be an effective means of producing events of type *E* at earlier times?

Similarly, suppose that an event of type *E* has occurred at location s_1 at time t_1. How does it follow that the later production of an event of type *C* cannot be an effective means of bringing about the occurrence of the earlier event of type *E*? For might it not be true that, if there *were not* an event of type *C* at location s_2 at time t_2, there *would not* have been an event of type *E* at location s_1 at time t_1? And, if that counterfactual were true, would not that be a reason for holding that the production of events of type *C* at later times is an effective means of producing events of type *E* at earlier times?

The response, in short, is that it has not been shown in the above argument that certain crucial counterfactuals are false - counterfactuals whose truth would support the claim that bringing about events of type *C* at later times *is* an effective means of bringing about events of type *E* at earlier times.

Can the argument be supplemented to meet this objection? One possibility is to appeal to the idea that 'backtracking' is not to be allowed in the case of standard counterfactuals - that is, in evaluating a counterfactual of the form 'If *C* were the case, then *E* would be the case', one keeps fixed all states of affairs up to the time of event *C* - since, if this procedure is followed, and one is in a world where an event of type *E* occurred at time t_1, and an event of type *C* at a later time t_2, then it will not be true that, if an event of type C had not occurred at time t_2, then an event of type *E* would not have occurred at the earlier time t_1.

The problem with this response is that it can be argued that, in interpreting standard counterfactuals, the correct principle concerning which events can be changed should be formulated in *causal* terms, rather than *temporal* ones. In particular, in evaluating a counterfactual of the form 'If *C* were the case, then *E* would be the case', it is not necessarily inadmissible not to keep fixed all of the states of affairs that are temporally prior to event *C*. What is essential, rather, is that change be confined to events that are not causally prior to *C*. In short, it is not *temporal* backtracking *per se* that must be excluded, but *causal* backtracking.

In the absence of backward causation these may come, of course, to the same thing, but, if one assumes, for example, that time travel into the past is possible, they often will not, and in such cases it is causal backtracking, not temporal backtracking, that must be excluded in evaluating counterfactuals, since surely one wants to say that, had Dr. Who not been wearing a red scarf when he stepped into Tardis in AD 2050, then neither would he when he emerged thirty years earlier, in AD 2020

This latter claim might be disputed, but only, I think, by arguing that, in evaluating a counterfactual, the events that should, so far as possible, be kept fixed are those that are *objectively fixed*, as of the time of the state of affairs referred to in the antecedent. Then, in the case of the counterfactual 'If Dr. Who had not been wearing a red scarf when he stepped into Tardis in AD 2050, neither would he have had one when he emerged in AD 2020', the events that are to be kept fixed will be those that - aside from Dr. Who's wearing a red scarf - are objectively fixed as of the relevant time in AD 2050, and, it might be contended, these will include Dr. Who's having a red scarf when he emerges from Tardis in AD 2020 So the counterfactual in question will be false: even if Dr. Who had not been wearing a red scarf when he entered Tardis in AD 2050, he would still have had one when he emerged in AD 2020

But, if the advocate of the argument from causation and potential control adopts this line of defense, then he or she is no longer advancing an argument that is neutral with regard to the nature of time: a tensed view is being presupposed, and one that entails that the future, unlike the past, is not fixed. Accordingly, this argument is, in the end, open to the same objection as the preceding argument.

13.3 Four Neutral Arguments against Backward Causation

The other four arguments that I shall consider, by contrast, are neutral between tensed and tenseless accounts of the nature of time. I shall argue, however, that all of them run afoul of a certain objection concerning the relation between backward causation and causal loops.

13.3.1 Backward Causation and the Bringing-About of Contradictions

The first of these arguments appeals to the idea that, if backward causation were logically possible, then it would be possible to make contradictions true. This argument is illustrated, for example, by the time travel case discussed by David Lewis:

> Consider Tim. He detests his grandfather, whose success in the munitions trade built the family fortune that paid for Tim's time machine. Tim would like nothing so much as to kill Grandfather, but alas he is too late. Grandfather died in his bed in 1957, while Tim was a young boy. But when Tim has built his time machine and traveled to 1920, suddenly he realizes that he is not too late after all. He buys a rifle; he spends long hours in target practice; he shadows Grandfather to learn of his daily walk to the munitions works; he rents a room along the route; and there he lurks, one winter day in 1921, rifle loaded, hate in his heart, as Grandfather walks closer, closer, . . .[3]

It is natural to suppose that it would, in such a situation, be possible for Tim to kill Grandfather. But the course of history leading up to the existence of Tim does not contain any state of affairs which is Grandfather's being killed in 1921. So, if Tim succeeds in killing Grandfather, then it is both true, and not true, that Grandfather is killed in 1921. It would seem, therefore, that if backward causation were possible, then it would also be possible for contradictions to be true. Consequently, backward causation must be impossible.

13.3.2 Backward Causation and the 'Undercutting' of Causal Chains

Another familiar argument for the view that backward causation is logically impossible is the 'bilking' argument, advanced by Antony Flew, Max Black, and others,[4] and which might be put as follows. Consider the hypothesis that later events, of type *L*, are causally necessary for earlier events, of type *E*. Confronted with such a causal hypothesis, Flew argues, one should not be satisfied with the role of a passive observer; one should not merely stand by to see whether, as it happens, earlier events of type *E* are always followed by later events of type *L*. Rather, one should subject such a causal hypothesis to vigorous experimental investigation. And one way of doing that is to wait until an event of type *E* occurs, and then try to prevent the occurrence of any appropriately related event of type *L*. If one does this, there are two possible outcomes. On the one hand, one may sometimes succeed. If so, the hypothesis has been falsified: events of type *L* are not causally necessary for events of type *E*. On the other hand, suppose that, when an earlier event of type *E* has occurred, one never

---

[3]David Lewis, 'The Paradoxes of Time Travel', *American Philosophical Quarterly*, 13 (1976), 145-52. See p. 149.

[4]See, for example, Antony Flew, 'Can an Effect Precede its Cause?', *Proceedings of the Aristotelian Society*, Suppl. 28 (1954), 45-62, and Max Black, 'Why Cannot an Effect Precede its Cause?', *Analysis*, 16 (1955-6), 49-58.

succeeds in preventing the occurrence of an appropriately related later event of type *L*. Then the question to ask is what makes it *impossible*, given an earlier occurrence of an event of type *E*, to prevent the occurrence of an event of type *L*? The only satisfactory answer, Flew and others seem to say,[5] is that the causal antecedents of later events of type *L* are not in fact independent of earlier events of type *E*, so that, rather than events of type *L* being causally necessary for events of type *E*, the latter are causally sufficient for the former. The unbreakable correlation is to be explained via forward causation, rather than backward causation. So there could never be a situation in which belief in backward causation would be justified.[6]

13.3.3 Causation and Increase in Probability

One of the most carefully developed arguments for the claim that backward causation is logically impossible is that advanced by Hugh Mellor in his book *Real Time*.[7] As Mellor points out, his argument is an extension of Michael Dummett's discussion in the latter's 'Bringing about the Past'.[8] Mellor's argument also resembles, with its central focus upon the experimental evaluation of causal claims, the argument considered in the preceding section. But, as we shall see, Mellor's argument involves the introduction of other crucial elements - and the result is a much more tightly constructed and circumspect argument.

Mellor's argument has two main parts. First, he argues that backward causation entails the existence of causal loops. Secondly, he attempts to show that causal loops are logically impossible. If both claims are correct, then backward causation is logically impossible.

Let us consider, then, the first step. Why could there not be a world that contained backward causation, but no causal loops? Mellor appeals, at this point, to a view that he advanced in an earlier chapter, to the effect that events

---

[5]See, for example, Antony Flew, 'Effects before their Causes? - Addenda and Corrigenda' *Analysis* 16 (1955-6), 104-10, especially p. 107.

[6]Critical discussion of this argument can be found in Michael Scriven, 'Randomness and the Causal Order', *Analysis*, 17 (1956-7), 5-9.

[7]D. H. Mellor, *Real Time* (Cambridge: Cambridge University Press, 1981), ch. 10.

[8]Michael Dummett, 'Bringing about the Past', *The Philosophical Review*, 73 (1964), 338-59, and reprinted in Michael Dummett, *Truth and Other Enigmas* (Cambridge, Massachusetts: Harvard University Press, 1978), 333-50.

always have spatiotemporally contiguous effects.[9] Given this premise, if a later event *L* causes an earlier event *E*, then *E* will necessarily have later effects in its immediate neighborhood, and those events in turn will have later effects, until eventually *E* gives rise to events that occur at the same time as event *L*. Even if this is so, however, it would not by itself seem to entail that *E* is among the causal antecedents of *L*. But perhaps there are ways in which this gap can be bridged. Alternatively, perhaps Mellor really wants to defend a weaker thesis than what I have attributed to him - to the effect that backward causation, even if it need not entail the actual existence of causal loops, must at least entail that causal loops are physically possible. This more modest claim, together with the proposition that causal loops are logically impossible, will still entail the impossibility of any backward causation, and the more modest claim could be defended by appealing to the idea that, if all events must have spatiotemporally contiguous effects, then there could always be circumstances that would result in a forward causal chain that would intersect the backward causal chain in question, thereby resulting in a causal loop.

Let us turn, then, to the second part of Mellor's argument. Why are causal loops conceptually problematic? Mellor's argument here appeals to a claim concerning the relation between causation and probability - to wit, that 'causes have *inter alia* to make their effects more likely than in the circumstances they would otherwise have been'.[10] Given this assumption, his argument runs as follows. Imagine a world that is supposed to contain backward causation. In particular, suppose that striking of matches can, in appropriate circumstances, cause earlier ignition of fires. Suppose further that such fires do give rise to later effects in their immediate vicinity, which in turn give rise to later effects, and so on, so that causal loops involving the backward causal chains are possible. It would then be in principle possible not only to perceive the fires that ignite, but to do so prior to the later acts of striking matches, and one could then use that information to control when matches are and are not struck. In particular, Mellor suggests that one can set up the following test: 'We can see by t_1 which fires ignited at t_0 and which did not: assume about equal numbers of each. We now ask half the self-styled igniters of each group to put matches to their fires at t_2 and the rest to refrain. They try to comply (or we should have no test), we note at t_2 whether they do, and then we assess the results.'[11]

---

[9]D. H. Mellor, *Real Time* (Cambridge: Cambridge University Press, 1981), 123.

[10]Ibid. 178.

[11]Ibid. 178.

What are the possible outcomes? One is that there is a positive correlation between the striking of matches and the earlier lighting of fires because some of the igniters disobeyed the instructions. Given that result, the natural idea, as Mellor points out, is to rerun the experiment, offering more glorious rewards to igniters who obey, and more exciting punishments to those who do not. But what if a correlation continues to exist, despite all efforts to improve the behavior of the igniters? Then, Mellor says, the appropriate conclusion to draw is that there is a causal connection between the striking of matches and the earlier ignition of fires, but it is not one that runs in the backward direction: 'Somehow fires igniting at t_0 make their igniters more likely to apply matches at t_2 than unlit fires do. People's activities at t_2 are somehow constrained, albeit only statistically, by what happens to fires at t_0.'[12]

A second possible outcome is that there is a positive correlation between the later striking of matches and earlier ignitions, and this is so even though all of the igniters in both groups obey their instructions. But how could there be a correlation in that case? The only possibility is that mistakes were initially made concerning which fires were lit, and which were not. But it would seem that it should be possible to reduce such mistakes to a very low level by the use of a sufficient number of accurate perceivers and recording devices.[13] So it would seem that one could also exclude this second possible outcome.

But the only other possibility is that the earlier perceptions of whether fires have lit are accurate, and that the igniters obey their instructions - in which case the result must be that the ignition of fires does not exhibit any positive correlation with later strikings of matches. It then follows, in virtue of the relation between causation and increase in probability, that strikings of matches cannot cause earlier fires to ignite. Backward causation is, accordingly, impossible.

13.3.4 A Formal Approach: Causation and the 'Transmission' of Probabilities from Cause to Effect

The final approach to the question of backward causation that I want to mention involves the attempt to formulate postulates that can plausibly be viewed as expressing conceptual truths about causation, and that will entail that causal loops are logically impossible.

---

[12]Ibid. 178-9.

[13]Mellor advances, at this point, a more complicated response that involves consideration of the nature of perception, and its relation to prediction.

Merely finding postulates that have such consequences is not, of course, especially difficult. Thus, for example, if the postulates include ones asserting that causation is irreflexive and transitive, then it follows immediately that causal loops are impossible. But the very immediacy of the entailment means that such postulates are question-begging with respect to whether causal loops are possible. What is needed are postulates that will recommend themselves both to those who think that causal loops are possible, and to those who do not, but which in fact entail - in a relatively deep, and unobvious way - the impossibility of causal loops.

Let me now try to motivate a set of postulates that will satisfy these conditions. The postulates that I shall set out are concerned with causal laws, rather than with the relation of causation itself. This is not, of course, any limitation if one holds, as most philosophers do, that causal relations are logically supervenient upon laws plus the non-causal properties of, and relations among, things. But what if, instead, one embraces a singularist conception of causation, and holds that events can be causally related without that being an instance of any causal law at all? Then it might well seem mistaken to start out from postulates that deal with causal laws, rather than with causation. Elsewhere I have argued, however, that this is not so.[14] First of all, it certainly seems possible to hold that the concept of the relation of causation is to be analyzed in terms of the concept of causal laws, without being compelled thereby to reject a singularist approach to causation. The reason is that one can characterize causation as that relation such that, *if* it enters into laws, those laws must necessarily be causal laws - and where the latter are characterized by the postulates to be set out below. Secondly, it seems to me that there are strong reasons for holding, first, that causation is not an observable relation; secondly, that no reductionist account of causation can be satisfactory; and, thirdly, that no theory that does not refer to causal laws can serve to characterize uniquely the relation of causation. If these three theses are correct, then there would seem to be no alternative to an approach that characterizes causal relations indirectly in terms of causal laws.

Let us ask, then, what it is that makes something a *causal* law. Reflection upon certain simplified situations suggests, I believe, a very plausible answer. Imagine, for example, the following possible world. It contains two radioactive elements, *P* and *Q*, that, in every sort of situation but one, exhibit half-lives of five minutes and ten minutes respectively. However, in one special sort of

---

[14]Michael Tooley, 'The Nature of Causation: A Singularist Account', *Canadian Journal of Philosophy*, Suppl. 16 (1990), 271-322.

environment - characterized by some property R - an atom of type P undergoes radioactive decay *when and only when* one of type Q also decays. Now, given such facts, a natural hypothesis would be that the events in question are causally connected. Either (1) the decay of an atom of type P is, in the presence of property R, both causally sufficient and causally necessary for the decay of an atom of type Q, or (2) the decay of an atom of type Q, in the presence of property R, is both causally sufficient and causally necessary for the decay of an atom of type P, or (3) there is some property S which is, given property R, causally sufficient and causally necessary both for the decay of an atom of type P and for the decay of one of type Q. But which of these causal connections obtains? Given only the above information, there is no reason for preferring one causal hypothesis to the others. But suppose that the following is also the case: in the presence of property R, both element P and element Q have half-lives of five minutes. Then surely one has good grounds for thinking that, given the presence of property R, it is the decay of an atom of type P that is both causally sufficient and causally necessary for the decay of an atom of type Q, rather than vice versa, and rather than there being some other property that is involved. Conversely, if it turned out that, in the special situation, both elements had half-lives of ten minutes, one would have good grounds for concluding instead that it was the decay of an atom of type Q that, given the presence of property R, was both causally sufficient and causally necessary for the decay of an atom of type P.

In the former case, the observed facts suggest that atoms of type P have the same probability of decaying in a given time period in the special situation that they have in all other situations, while atoms of type Q do not. On the contrary, atoms of type Q appear to have, in the special situation, precisely the same probability of decay as atoms of type P. So the probability of decay has, so to speak, been transferred from atoms of type P to atoms of type Q. It is this, I suggest, that makes it natural to say, in that case, that it is the decay of atoms of type P that is, in the presence of property R, both causally sufficient and causally necessary for the decay of atoms of type Q. The direction of causation coincides, therefore, with the direction of transmission of probabilities.

The basic idea, accordingly, is to characterize causal laws in terms of the transmission of probabilities. Talk about the 'transmission of probabilities' is, of course, a metaphor, and one needs to show that that metaphor can be cashed out in precise terms. How can this be done?

In order to set out in a perspicuous fashion the postulates that provide, I believe, the answer to this question, it will be helpful to introduce a few symbols. First, I shall use the expression '$P \Rightarrow Q$' to say that it is a causal law that the possession of property P is causally sufficient to bring about the possession of property Q. Secondly, I need to refer to relations of logical probability. To do this, I shall use the expression 'Prob(Px, E) = k', where this is to be interpreted as saying that the logical probability that x has property P, given only evidence E, is equal to k. Finally, I need to refer to information of a certain restricted sort - specifically, information that either is tautological, or that concerns only what causal laws there are. I shall use the term C for that purpose.

Given the above notation, one natural formulation of the desired postulates for causal laws is this:

(C_1): $\text{Prob}(Px, P \Rightarrow Q \ \& \ C) = \text{Prob}(Px, C)$

(C_2): $\text{Prob}(Qx, P \Rightarrow Q \ \& \ C) = \text{Prob}(Px, C) + \text{Prob}(\sim Px, C) \times \text{Prob}(Qx, \sim Px \ \& \ P \Rightarrow Q \ \& \ C)$

(C_3): $\text{Prob}(Qx, \sim Px \ \& \ P \Rightarrow Q \ \& \ C) = \text{Prob}(Qx, \sim Px \ \& \ C)$

(C_4): $\text{Prob}(Qx, P \Rightarrow Q \ \& \ C) = \text{Prob}(Px, C) + \text{Prob}(\sim Px, C) \times \text{Prob}(Qx, \sim Px \ \& \ C)$.

These postulates are essentially somewhat simplified versions of ones that I set out elsewhere in developing a supervenience account of causation.[15] There I discussed, in a fairly detailed way, the line of thinking that leads to the specific postulates in question.[16] So perhaps it will suffice here simply to note the central considerations.

Postulate (C_1) states that, if the prior probability that some individual will have property P, given only information that is restricted to logical truths and statements of causal laws, has a certain value, then the posterior probability of that individual's having property P, given the additional information that the possession of property P causally gives rise to the possession of property Q, must have precisely the same value. Postulate (C_1) therefore asserts, in effect, that the posterior probability of a state of affairs of a given type is, in the situation described, not a function of the prior probability of any state of affairs of a type to which states of affairs of the first type causally give rise.

Postulates (C_2), (C_3), and (C_4) deal with the posterior probability of a state of affairs, given information to the effect that it is a state of affairs of a type that is causally brought about by states of affairs of some other type, together with prior

---

[15]It is also possible to set out corresponding postulates for probabilistic causal laws, but they are unnecessary in the present context. (For details concerning the probabilistic analogues, see my *Causation: A Realist Approach* (Oxford: Oxford University Press, 1987), 291-6.)

[16]Ibid. 256-62.

information that is restricted in the way indicated above. The first of these three postulates asserts that, given the additional information that the possession of property *P* causally gives rise to the possession of property *Q*, the posterior probability that some individual has property *Q* is, in the way indicated, a function of the prior probability that that individual has property *P*.

Postulate (*C*2) does not, however, express that dependence in the clearest way, since it involves, on the right-hand side, a probability that is also conditional upon the information that the possession of property *P* gives rise to the possession of property *Q*. It is for this reason that postulate (*C*3) is part of the theory, for it makes it possible to derive a statement in which the relevant posterior probability is expressed in terms of prior probabilities alone.

Postulate (*C*3) asserts that, if the prior probability that some individual will have property *Q*, given only information that is either tautologous, or else restricted to statements of causal laws, has a certain value, then the posterior probability of that individual's having property *Q* - given the additional information both that the possession of property *P* causally gives rise to the possession of property *Q*, and that the individual does *not* have property *P* - must have precisely the same value as the prior probability. (*C*3), together with (*C*2), then entails (*C*4), which does express the posterior probability that an individual will have property *Q* in a way that involves only prior probabilities.

The crucial content of the above theory of causal laws is expressed, accordingly, by postulates (*C*1) and (*C*4), since (*C*1) expresses the fact that the logical probability that a given state of affairs will obtain, given information about the types of states of affairs that *are caused by* states of affairs of that type, does not differ from the prior probability, upon evidence of a certain restricted sort, that the state of affairs in question will obtain. Its posterior logical probability cannot, therefore, be a function of the prior logical probabilities of states to which it causally gives rise. But by contrast, as is indicated by postulate (*C*4), the posterior logical probability of a given state of affairs is a function of the prior logical probability of any state of affairs of such a type that states of affairs of that type *causally give rise to* states of affairs of the first type. The relation between posterior probabilities and prior probabilities is, in short, different for causes than for effects.

Do the above postulates capture the intuition that there is a relation between causation and the direction of transmission of probabilities? I believe that they do. For if one applies them to the example above, it turns out that, if the radioactive decay of an atom of type *P* in the presence of property *R* is both causally necessary and causally sufficient for the decay of an atom of type *Q*, then postulate C_1 entails that the posterior probability of an atom of type *P*'s decaying in a given period of time is equal to the prior probability of that sort of event, whereas postulate C_4 entails that the posterior probability of an atom of type *Q*'s decaying in a given period of time, rather than being equal to the prior probability of that sort of event, is equal instead to the prior probability of an atom of type *P*'s decaying. So the postulates generate the correct probability values.

One way of motivating, therefore, the adoption of postulates (C1) through (C4) is by appealing to the intuition, elicited by examples such as that used above, that there is a connection between the direction of causation and what I have referred to as the direction of transmission of probabilities. But the case for the above postulates does not rest upon that intuition alone: there are at least three other grounds of support that can be offered. The first is that reductionist accounts of the direction of causation are open to decisive objections - objections that an account based on postulates (C1) through (C4) totally avoids. The second is that those postulates - or, rather, slightly strengthened versions of them - generate the formal properties of causal relations. The third is that the above postulates also serve to explain how beliefs about causal relations can be justified.[17]

Do the above postulates have any implications with respect to the possibility of causal loops? The answer is that quite strong conclusions can be derived from them. First, causal loops where each event is causally sufficient for the next event in the loop are logically impossible. Secondly, the same is true with respect to loops in which each event is causally necessary for the next event. Thirdly, let us say that one event is causally prior to another if and only if it is either causally sufficient, or causally necessary, for the other. Then it is logically impossible to have loops in which each event is causally prior to the next event.

These results, moreover, are not obvious or trivial consequences of the postulates: the derivations are in fact rather complicated.[18] So the postulates cannot be described as begging the question with respect to the possibility of causal loops. Nevertheless, someone who thought that causal loops of one or more of the sorts just mentioned were in fact possible would - assuming the

---

[17]For a defence of the first claim, see my 'Causation: Reductionism versus Realism', *Philosophy & Phenomenological Research*, Suppl. 50 (1990), 215-36. Arguments in support of the second and third claims can be found in *Causation: A Realist Approach* (Oxford: Oxford University Press, 1987), 278-87 and 296-303, respectively.

[18]The relevant proofs are set out in *Causation*, 275-87 and 328-35.

derivations are sound - probably want to run the arguments in the opposite direction, and conclude that at least one of the postulates must be mistaken. However, given that the postulates do seem to capture plausible intuitions concerning the relation between causation and probability, and that they can be supported in the other ways mentioned above, it would seem unlikely that there is nothing in postulates (C_1) through (C_4). If so, then the crucial question would seem to be whether there is any way of weakening that set of postulates to get an alternative set that will have the same merits, but which will not entail the impossibility of certain sorts of causal loops. If that could be done, then the present argument would be completely undercut.

Suppose that no such weakening exists, that no set of postulates that does not entail the impossibility of causal loops can generate the right consequences concerning the relation between causation and probability, and provide both a non-reductionist account of the direction of causation, and an explanation of the possibility of causal knowledge. Where would that leave one? The answer is that one would still be left with the question of whether backward causation is possible - unless it is true, as Mellor and others have maintained, that backward causation entails at least the possibility of causal loops. In the next section, however, we shall see that there are reasons for holding that backward causation does not entail the possibility of causal loops. So, even if one can show that causal loops are impossible by appealing to postulates (C_1) through (C_4), that will not suffice to establish the impossibility of backward causation.

13.4 The Basic Difficulty: Backward Causation without Causal Loops

The four arguments just considered are all directed to showing, in the first instance, that causal loops are impossible. The question therefore arises as to whether one can move on to the further conclusion that backward causation is impossible. To do so, one will somehow have to prove that backward causation entails the possibility of causal loops. But can that be done?

Given certain views concerning the independence of causation and time, such a line of argument would be impossible. In particular, suppose that the direction of time was not based upon the direction of causation. Then among the worlds that would have to be regarded as possible, until proven otherwise, would be worlds that contained backward causation, but where forward causation was not only completely absent, but impossible. So there would be possible worlds containing backward causation, but without even the possibility of causal loops, and the above line of argument would therefore collapse.

Some writers have been aware of this point. Hugh Mellor, in particular, emphasizes the role that is played in his own argument by a causal account of the direction of time. But many other philosophers have failed to consider the possibility of a world that contains only backward causation, even when it seems that they do not accept a causal analysis of the direction of time.

I shall not appeal, however, to the idea of a world that contains only backward causation. Like Mellor, I believe that causation fixes the direction of time, and thus that it is not really possible for there to be a world containing only backward causation. What I shall need to argue, accordingly, is that there can be worlds where, in spite of the presence of both backward and forward causation, causal loops are impossible.

Consider, then, the following world. It consists of two spatial regions segregated from each other by a wall with some remarkable properties. First, the wall has always existed. Secondly, it is indestructible, and so it will exist at every future time. Thirdly, there is no way of getting around the wall, so that something can travel from the one region of space to the other only by going through the wall. Finally, the wall's properties differ depending upon the direction through the wall: in one direction, no causal processes at all can be transmitted from the one region to the other, while, in the opposite direction, light waves, but nothing else, can pass through the wall.

Suppose now that you are living on the side of the wall that can receive light waves from the other side, but from which no causal influence at all can be transmitted back in the opposite direction. On your side of the wall, life is very similar to the actual world. But on the other side, things are somewhat different. For example, next to you, on the other side, is a golf course. As you look across, you see a golf ball, lying on a green. But then, though there are no readily observable forces acting upon the ball, it starts to roll, at first slowly, and then more rapidly. It then hops, landing exactly in an indentation on the green, out of which it leaps at a much higher velocity. You notice that, remarkably, the indentation that was there has now completely disappeared. Following the flight of the ball, you notice it moving towards a golfer who appears to be swinging a club in reverse. As the ball moves towards the ground, you also notice that a loose piece of turf, that had been lying there, leaps in the air, and the turf, the golf ball, and the clubhead all arrive at the same location at about the same time. The piece of turf winds up fitting exactly into a divot hole that had been there in front of the golfer, while the ball comes to a complete halt as it makes contact with the backward-moving clubface.

A remarkable series of events. But in the world we are imagining, these sorts of events take place all the time on the other side of the wall: it is as if all causal processes were running in the opposite direction from what they are on this side of the wall. So let us assume now that things are just as they appear to be. Causal processes on the other side run in the opposite direction from those on this side.

We have, then, a world with oppositely directed causal processes, but where causal loops are not even possible. Seeing the golf course, and the peculiar behavior of the ball on the green, you may very well be able to predict what the golfer standing down the fairway is going to do. But if you attempt to subject the apparent causal connection to any experimental test, you will necessarily fail, since no causal processes can pass from your side to the other. Experimenters on the other side, of course, will have no trouble subjecting the relevant causal hypothesis to an experimental test, and in finding, thereby, that golfers on their side, by swinging a golf club, can indeed control - at least to some extent - the flight of a golf ball.

In this first type of world, events belonging to oppositely directed causal chains on different sides of the wall can be causally connected. The golfer's swing causes the ball to be on the green, which then causes your perceptual experiences, which, in turn, cause your later memories. Then, a few seconds later, the golfer's swing causes your perceptual experiences of it, and those experiences together with your earlier memory produce a memory that encompasses both the swing and the outcome. So in a certain sense one has events that form a causally interrelated loop. But the sort of causal interrelatedness is harmless, and does not constitute a causal loop, since, as one goes around the loop, the connecting relation at one point, rather than being that of causation, is instead the inverse relation of being caused by.

The second type of world, by contrast, is one where the oppositely directed causal processes are not causally connected in this way - nor, indeed, in any way at all. This second possibility can be arrived at by a two-step modification of the segregated world. The first modification involves imagining that there are different types of fundamental particles on different sides of the wall, whose natures are such that there would be no direct causal interaction between them even if the wall were not present. So now it cannot be the case that light is being transmitted through the wall in one direction. What must be happening is that light of the type found on the other side interacts with the wall, and produces a causal process in the wall, which leads in turn to the emission of light of another type on this side of the wall. The existence of the wall would then be necessary for any causal interaction between things on different sides of the wall.

The second modification then consists of removing the wall. The two parts of the resulting world will then be causally isolated from one another. They will, however, still be spatially related. Thus, though you will no longer be able to see the ball lying on the green in the sub-world made up of different types of particles, that state of affairs will be spatially related to events in your sub-world, and some of those events may very well cause events that will be spatially related to the golfer's swing. So one can have oppositely directed causal processes entering into loops of the following sort: C_1 causes E_1; E_1 is spatially related to C_2; C_2 causes E_2; E_2 is spatially related to C_1. Indeed, given that there is now no wall to keep the two sub-worlds apart, we might as well imagine that, rather than occupying different regions of space, they overlap completely. Then one could have oppositely directed causal processes entering into loops of the following sort: C_1 causes E_1; E_1 and C_2 occur in precisely the same spatiotemporal location; C_2 causes E_2; E_2 and C_1 occur in precisely the same spatiotemporal location.

In short, it seems possible for there to be oppositely directed causal processes in worlds where causal loops are impossible. This may happen either because oppositely directed causal processes, though spatially related, cannot be causally connected, or because, though oppositely directed causal processes can be causally connected, there are restrictions upon the possible causal connections that are built into the world, and that preclude causal loops.

Do we have here a decisive objection to any attempt to rule out backward causation by means of an argument that presupposes a connection between backward causation and the possibility of causal loops? I think that would be to overstate the matter, since there are ways in which one might attempt to show that worlds of the sort to which I have just appealed are not in fact really possible. In particular, one might try to argue that spatiotemporal relations need to be analyzed in causal terms, and, if that could be established, then the present objection might well be undermined.

But how promising is such a response? It seems to me that the prospects are not bright, since if one considers what is undoubtedly the most impressive attempt to analyze spatiotemporal relations in causal terms - namely, Alfred A. Robb's axiomatic development of the space-time of the Special Theory of Relativity[19] - a problem emerges that seems likely to plague all such attempts. The starting-point of the difficulty is that there seems to be no reason why *actual* causal relations among events will necessarily suffice to fix the total structure of space-time. Because of this fact, Robb's own account is formulated not in terms of actual causal relations between events, but in terms of causal connectibility. The concept of causal connectibility, however, presumably has to be analyzed in

---

[19]Alfred A. Robb, *A Theory of Time and Space* (Cambridge: Cambridge University Press, 1914).

terms of counterfactuals, and so one is faced with the question of the truth-makers for the relevant subjunctive conditionals: what makes it true that, although event *A* is not causally connected to event *B*, there could have been a causal process connecting those events, or, at least, connecting alternative events in the same spatiotemporal locations? It is hard to see what answer there can be, other than one that appeals not only to relevant laws of nature, and to the intrinsic properties of the events in question, but also to the spatiotemporal relation between those events. This answer is not available, however, if one holds that spatiotemporal relations are themselves to be analyzed in terms of actual and possible causal relations.

The problem, in short, is that the most natural way of attempting to answer the argument developed in this section seems to require subjunctive conditional statements for which, on pain of circularity, no categorical truth-makers can be supplied. Perhaps this obstacle can be overcome. Or perhaps there is some alternative way of attempting to show that worlds of the sort described above are not really possible. But until a promising proposal appears on the horizon, I think that the conclusion must be that backward causation need not entail the possibility of causal loops, and thus that any attempt to rule out backward causation which claims that there is such an entailment must necessarily fail.