CAUSATION

Chapter 1

Causation: Historical Background, Basic Issues, and Alternative Views

When did philosophers first realize that the idea of causation was a philosophically challenging and potentially problematic concept? After considering this question, I shall go on, in this introduction, to set out the most important philosophical questions that have since arisen in connection with causation – questions which any satisfactory account of the nature of causation must answer. Then I shall outline the most basic divisions into which approaches to the philosophy of causation fall, so that readers have, from the beginning, a good idea of the main alternatives. Finally, I shall briefly describe two important underlying issues with which it is crucial to grapple if one is to arrive at defensible answers to the fundamental philosophical questions concerning causation.

1. David Hume and the Discovery of the Problem Posed by Causation

Causal concepts have surely been present from the time that language began, since the vast majority of action verbs involve the idea of causally affecting something. Thus, in the case of transitive verbs of action, there is the idea of causally affecting something external to one – one finds food, builds a shelter, sows seed, catches fish, and so on – while in the case of intransitive verbs describing physical actions, there is the idea of causally affecting one's own body – as one walks, runs, jumps, hunts, and so on.

It was not long after the very beginning of philosophy in ancient Greece that serious reflection concerning the nature of causation arose – with Aristotle's famous discussion of causation in Book 2 of his *Physics*. The result was Aristotle's doctrine of four types (or, perhaps, aspects) of causes – material, formal, efficient, and final.

Consider, for example, the construction of a house. The material aspect of the cause would consists of the materials – wood, brick, cements, and so on – used in the construction of the house. The formal aspect of the cause would be the way that those materials would be arranged in the completed house. The efficient cause would consist of the activities of the builder who brought about that arrangement of the materials. The final cause would be the plan in the mind of the builder that directed the builder's construction activity.

Aristotle's account that was immensely influential for about 2000 years. What was not realized, however, either by Aristotle, or by any other philosopher

during that stretch of time – perhaps because of the sense of familiarity with the idea of causation occasioned by the almost ubiquitous presence of causal concepts in even the most rudimentary parts of language – is that the concept of causation gives rise to very serious, puzzling, and difficult philosophical questions. Thus it was only many centuries after Aristotle (384-322 BCE), with the appearance of David Hume (1711-1776) and his famous discussions of the relation of cause and effect (1739-40 and 1748), that philosophers realized that the idea of causation was by no means simple and straightforward.

Why did Hume see what so many thoughtful philosophers before him had not? The reason is that Hume held – as did the other British empiricists, Locke and Berkeley – that while some descriptive, non-logical concepts can be analyzed in terms of other descriptive concepts, in the end analysis of all such concepts must terminate in ideas that apply to things in virtue of objects' having properties and standing in relations that can be immediately given in experience. Hume therefore asked whether the relation of causation was one that could be given in immediate experience, and he argued that it could not. The question for Hume, accordingly, was how the concept of causation could be analyzed in terms of ideas that do pick out properties and relations that are given in experience, and once this question was in view, Hume was able to show that arriving at a satisfactory answer was a very difficult matter.

2. Causation: The Most Basic Philosophical Questions

The fundamental questions that must be answered by any adequate theory of causation fall into three main groups. First, there are issues concerned with the very idea of causation, with the concept of causation itself. Secondly, once one is clear about the concept of causation, there are epistemological questions concerning the justification of beliefs about causal states of affairs. Thirdly, there are questions about what are called the formal properties of causation. Finally, there are questions about the relations between causation on the one hand, and space and time on the other. So let us consider each of these in turn.

2.1 Causal Relations, Causal Laws, and the Concept of Causation

Among our causal beliefs are ones affirming that two particular events, or states of affairs, were causally related: the ball's hitting the window caused the window to break. As mentioned above, the presence of transitive verbs of a very basic sort in all human languages testifies to the presence of such beliefs from very early times. But such beliefs about causal relations between events at particular times and places are surely connected with *general* causal beliefs, with beliefs that events of one type are causally related to events of some other type – such as that pushing against an object causes the object to move.

Such early general causal beliefs were undoubtedly rough and ready, and often subject to many exceptions. If one picks too large an object, for example, pushing on that object may very well not result in any movement. Over time, such causal generalizations could be refined, of course, by adding, for example, clauses that restricted the scope of the initial generalizations, thereby excluding cases where the causal connection did not obtain.

Eventually, causal generalizations of a very important sort were discovered: causal laws. The journey from the early causal generalizations that humans arrived at relatively easily to those causal generalizations that are laws of nature was, however, a very long and difficult one, and it took the genius of Isaac Newton, standing, as he said, on the shoulders of giants, to arrive, for example, at his Second Law of Motion – the law that if there is a total force *F* acting upon a body, that force causes the body to undergo an acceleration that is proportional to *F*, and inversely proportional to the mass of the body.

The upshot is that there are at least two types of causal propositions of which we need to be able to provide an account, namely, propositions concerning causal *relations* between particular, concrete states of affairs and events, and causal *laws* connecting types of events, or states of affairs. So we have the following questions:

(1) What it is for two particular events or states of affairs to be causal related?

(2) What are causal laws?

In addition, however, it would seem unlikely that these two types of causal states of affairs are not related in any way. Another question that needs to be considered, then, is this:

(3) How exactly are causal laws, on the one hand, and causal relations between states of affairs, on the other hand, related?

The various answers that can be offered will be considered later, but to flesh out the question a bit, here are two important possibilities. First, it could be that causal relations between events are the basic causal states of affairs, and that causal laws are simply regularities involving causal relations between events. On this view, what makes it a causal law that all events of type *C* cause events of type *E* is simply that every event of type *C* that exists stands in the relation of causation to some event of type *E*.

A second possibility is that it is causal laws that are the basic causal states of affairs, and that for a particular event, *c*, to cause an event, *e*, is just for there to be some causal law under which events *c* and *e* fall. But what does it mean to say that two events fall under some causal law? Here is one possibility. Suppose that event *c* is an event of type *F*, event *e* is an event of type *G*, that event *c* occurs at a temporal distance *d* before event *e*, and that it is a causal law that any event of type *F* causally gives raise to an event of type *G* that occurs later at a temporal

distance *d*. Then events *c* and *e* would fall under the causal law linking events of type *F* and events of type *G*.

Getting clear about how different types of causal states of affairs are related is, then, important. But a question that is, I think, even more important philosophically is this:

(4) How are causal states of affairs related to non-causal states of affairs?

To get a feeling for what is involved in this question, consider the answer that David Hume advanced **(Ref.)**. Hume began by focusing upon a single instance of some type of causal relation. Consider, in particular, the case of two billiard balls, *A* and *B*, the first moving, the other at rest. Billiard ball *A* collides with billiard ball *B*, thereby causing *B* to move. What properties or relations, other than that of causation, are present in such a case? Hume's answer was as follows. First, the motion of billiard ball *A* that causes *B* to move temporally precedes the motion of *B*. Secondly, there is no temporal gap between the motion of ball *A* and *B*'s beginning to move. Thirdly, neither is there any spatial gap: *A* continues to move right up to the point where it is in contact with *B*, at which point *B* begins to move.

At least one philosopher – namely, C. J. Ducasse – has held that, though some refinements are needed, the three relations just mentioned are essentially all that there is to causation. Hume, however, thought that there was more to causation. But what other properties or relations can we observe when one billiard ball causes another to move? Hume's answer was that in any particular case there are no other properties or relations that are relevant to the causal connection. Nevertheless, there is something else that is relevant, namely the fact that the particular case in question falls under a generalization, in the sense explained above. Hume arrived, thus, at the following analysis:

[Insert definition given in the *Treatise*.]

Hume's conclusion, accordingly, was that for one event to be a cause of another event is nothing more than for the first event to be earlier than the second, for there to be no gap, either temporal or spatial between the two events, and for the two events to be of types such that every event of the first type is always followed by an event of the second type with which it is both spatially and temporally contiguous. Hume also thought that neither the relation of temporal priority, nor the relations of spatial and temporal contiguity, involved causal relations. If Hume was right about this – we shall see later that he may not have been right in the case of temporal priority – then what Hume has offered is a *reductionist* account of causation, according to which a causal relation between states of affairs is nothing over and above the existence of certain non-causal relations between those events, and other pairs of events of relevant sorts.

Many other reductionist accounts of causation can and have been offered, that appeal to other sorts of non-causal states of affairs, and we shall be

considering these later. But there is, of course, a very different answer to the question of how causal states of affairs are related to non-causal ones – namely, that causal states of affairs are something over and above non-causal ones.

But what does this "something over and above" mean? What is it to be a *non-reductionist* with regard to causation?

This is something we shall need to discuss more carefully later. At bottom, however, what any non-reductionist view of causation claims is that there are logically possible ways the world might be that involve exactly all the same non-causal facts, but that differ causally in one or more ways. Imagine, then, that worlds W_1 and W_2 are two possible worlds that are completely indistinguishable with respect to all non-causal states of affairs. If x is any instantaneous state of affairs in world W_1, let x^* be the corresponding instantaneous state of affairs in world W_2. According to a non-reductionist approach to causation, then, if x and y are two instantaneous states of affairs in W_1, it could be the case that while x does not cause y, the corresponding two instantaneous states of affairs in W_2, namely, x^* and y^*, are such that x^* does cause y^*.

2.2 Knowledge of Causation

Many concepts are analyzable, but either we have no reason to think that there is anything in the actual world to which those concepts apply, or, more strongly, we have good reason for thinking that there are no such things. Concepts such as those of a unicorn, or of caloric fluid, or of witches, are perfectly clear and comprehensible, but we either have no reason for believing there are such things, or we have good reason for thinking that there are no such things. Could causation be like that?

Initially, skepticism concerning the existence of causal relations might strike one as quite extraordinary. After all, as was noted earlier, natural languages contain an enormous number of verbs, both transitive and intransitive, that involve the idea of a causal relation, either between states of one's body and states of other objects, or states of one's mind and states of one's body. Could it really be true that ordinary sentences containing such verbs are never true, that no one, or example, has ever thrown something, or raised an arm?

The answer is that, as extraordinary as this may initially seem, it is a genuine possibility, and one that, moreover, needs to be seriously addressed. The reason emerges if one considers the case of color. What did one believe, as a child, when one believed that ripe tomatoes were red, and unripe ones green? Did one believe what one believes now when one says that ripe tomatoes were red, and unripe ones green?

Upon reflection, I think it is clear that that is not the case. For did not one believe, when one was young, that redness and greenness qualitative properties that were right there on the surfaces of the objects in question. Physics, however, provides us with good reason for holding that there are no such qualitative properties on the surfaces of objects. All that exists there, and that is relevant to our color experiences, are the powers that objects have to absorb certain frequencies of light, and to emit others. These powers, in turn, have their basis in the properties of, and relations among, the atoms and molecules that lie at the surface of the object. So unless qualitative properties such as redness and greenness can be reduced to the fundamental particles of physics, and their quantitative properties and relations, the external objects that we perceive are not colored in the sense that we thought they were before we became acquainted with the scientific picture of the world. It is hard to see, however, how any such reduction of qualitative redness and greenness to the stuff of physics is possible. The conclusion, accordingly, is that the extremely natural beliefs about objects being colored that we had when we were young, and that virtually all humans had before the development of physics, are in all probability false.

Classifying objects by means of color terms is, of course, extremely useful. Because of this, when humans learned that qualitative color properties are not properties of external objects, they did not abandon color language. Instead, the meanings of the terms were systematically altered, so that most people now use color terms to refer to powers in objects to produce experiences that involve the relevant qualitative color properties in normal human observers under normal conditions. So we still say that ripe tomatoes are 'red', and unripe ones 'green', and that the things we see are 'colored', even though the beliefs that one once expressed by using that terms are false.

The question, then, is whether, given that the beliefs that humans naturally form about objects being colored are mistaken, the same thing could be true in the case of causation. But how could that be in the case of causation? The answer is the same as in the case of color. Thus, just as physics provides us with grounds for concluding that external objects do not have qualitative color properties, so some philosophers have argued that physics also provides us with good reason for thinking that the world does not contain any causal relations between events, or any causal laws. Why so? This is a subject that I shall turn to in chapter **(X)**. Basically, however, there are two main lines of argument. One is that the mathematical formulation of the laws of physics do not incorporate the concept of causation. The other is that, on the one hand, one thing that seems absolutely central to the idea of causation is that the relation of causation possesses a direction, whereas, on the other hand, with one minor and apparently irrelevant exception, none of the laws of physics introduce any sort of asymmetry – such as a temporal asymmetry between past and future events – that could serve as, or be related to, the direction of causation.

If this skeptical challenge can be met, what account can be offered of our knowledge of, or our justification for believing in, the existence of causal relations between events or states of affairs? The answer to this question will depend very heavily upon what account of the concept of causation is correct. At one extreme, there is the type of non-reductionist view of causation according to which the concept of causation is an analytically basic concept. Given this view, it will not be possible to have any justified beliefs about causal relations between events unless some such beliefs can be non-inferentially, or non-evidentially justified – that is, justified without being justified on the basis of other justified beliefs. But a number of philosophers, including Elizabeth Anscombe **(Ref.)**, David Armstrong **(Ref.)**, and Evan Fales **(Ref.)**, have argued that we do have precisely such non-inferential knowledge of, or non-inferentially justified beliefs about causation.

Suppose, for present purposes, that this claim can be sustained. It is clear that not all of the beliefs concerning causal relations that we normally take to be justified can be non-inferentially justified. For example, one is surely justified in believing that the cause of the erratic movement of small visible particles, such as pollen, in a liquid – that is, Brownian motion – is caused by collisions with invisible things – the liquid's molecules. Such causal beliefs are surely not non-inferentially justified. If justified, they must be inferentially justified on the basis of evidence.

That some causal beliefs can be inferentially justified if the concept of causation is analytically basic seems unproblematic. The reason is if one has non-inferentially justified beliefs about causally related events, where the members of the pairs of events of the same two types – say *C* and *E* – then one may be justified in believing that some other event of type *C* is causally related to an event of type *E*, even if one has not observed the events in question.

This extension of one's justified causal beliefs via generalization has limits, however, since the inductive method that involves generalizing from instances cannot serve to justify beliefs about unobservable entities. But there is nothing that prevents an advocate of the view that the concept of causation is analytically basic from appealing to a very different type of inductive method that we shall be considering later, and which is variously referred to as 'abduction', 'hypothetico-deductive method', 'the method of hypothesis', and 'inference to the best explanation'.

Suppose, on the other hand, that one adopts a reductionist approach to the concept of causation. In that case, the truthmakers for a given causal belief will be the truthmakers for the non-causal propositions that enter into the reductive account of the concept of causation, and as long as there is no problem about the justification of the relevant non-causal beliefs, the epistemology of causation will be straightforward.

Consider, for example, Hume's reductive account of the concept of causation. Provided that one can be justified in believing, that events *c* and *e* are both spatially contiguous and temporally contiguous, that *c* is earlier than *e*, and that there are types of events *F* and *G* such that *c* is an event of type *F*, *e* is an event of type *G*, and all events of type *F* are followed by events of type *G*, and can be justified in believing all of those things without appealing to any causal beliefs, then one will be justified in believing that event *C* caused event *E*.

Suppose, finally, that one adopts a non-reductionist account of the concept of causation. Some philosophers, such as Huw Price **(Ref.)** have contended that causal beliefs could not then be justified. This claim will be examined later, in chapter **(Ref.)**. On the face of it, however, it is unclear why there should be a problem. The reason is that theories in physics involve numerous concepts – such as those of mass, charge, quark, force, and so on – that are given a non-reductionist interpretation, and one is surely justified in believing that those theories are at least good approximations to the truth. So why should the situation be different in the case of causation when it is given a non-reductionist interpretation?

Showing that some causal beliefs can be justified is not, however, the only epistemological challenge. One should also attempt to establish, first, that the types of causal beliefs that we ordinarily take to be justified can be justified, and, secondly, that the types of evidence that we believe to be sufficient to justify causal beliefs of a given type are in fact sufficient.

An especially important case here arises from the social sciences, where highly sophisticated methods of arriving at causal hypotheses on the basis of complex statistical information are used. The philosophical issue thus posed is how statistical information is evidentially relevant to causal hypotheses. Some analyses of the concept of causation bring in probabilistic notions, and then it may be relatively straightforward to establish the evidential relevance of statistical information. But on other accounts of the concept of causation, the situation may be problematic. Consider, for example, the view that the concept of causation is analytically basic. How on this view is one to forge a connection between a causal hypothesis, say, that all events of type *C* causally give rise to events of type *E*, and the statistical information that all events of type *C* are followed by events of type *E*? But if one cannot forge such a connection, how can one ever be justified in holding that the proposition that all events of type *C* are followed by events of type *E* is evidence for the conclusion that all events of type *C* causally give rise to events of type *E*?

2.3 Formal Properties of Causation

Relations themselves can have various properties. Consider, for example, the *greater than* relation. It is surely a necessary truth that, for any numbers *x* and

y, if *x* is greater than *y*, then *y* is not greater than *x*. Relations that have this property are said to be *asymmetric*.

It is also a necessary truth, for any number *x*, *x* is not greater than itself. Relations that have this property are said to be irreflexive.

Finally, it is also a necessary truth that, for any numbers *x*, *y*, and *z* if *x* is greater than *y*, and *y* is greater than *x*, then *x* is greater than *z*. Relations that have this property are said to be *transitive*.

If a relation is asymmetric, it follows that it is also irreflexive. The converse, however, is not the case. Consider, for example, the relation of *being a sister of*. No one is her own sister, so the relation is irreflexive. But it may be true both that Clare is the sister of Sue, and that Sue is the sister of Clare, so the relation is not asymmetric.

Consider, next, the relation of *being equal to*. Necessarily, for any numbers *x* and *y*, if *x* is equal to *y*, then *y* is equal to *x*. Also, it necessarily true, for any number *x* that *x* is equal to *x*. Any relation that has the first of these properties is said to be *symmetric*, while any relation that has the second is *reflexive*.

There are many other formal properties of relations, but the above, which can be defined as follows, will suffice for our purposes here:

R is a *reflexive* relation = def. Necessarily, for all *x*, *x* stands in relation *R* to *x*.

R is an *irreflexive* relation = def. Necessarily, for all *x*, *x* does not stand in relation *R* to *x*.

R is a *symmetric* relation = def. Necessarily, for all *x* and *y*, if *x* stands in relation *R* to *y*, then *y* stands in relation *R* to *x*.

R is an *asymmetric* relation = def. Necessarily, for all *x* and *y*, if *x* stands in relation *R* to *y*, then *y* does not stand in relation *R* to *x*.

R is a *transitive* relation = def. Necessarily, for all *x*, *y*, and *z*, if *x* stands in relation *R* to *y*, and *y* stands in relation *R* to *z*, then *x* stands in relation *R* to *z*.

Two important questions that now arise in the case of causation are these:

(1) What are the formal properties of the relation of causation?

(2) How can one prove that causation has those formal properties?

As regards the first of these questions, the natural answers are as follows. First, it is logically impossible for any event or state of affairs to be the cause of itself, so causation is an irreflexive relation. Secondly, it is also logically impossible, if event *a* is the cause of event *b*, for event *b* to be the cause of event *a*, so causation is an asymmetric relation. Finally, it is logically necessary, event *a* is the cause of event *b*, and event *b* is the cause of event *c*, that event *a* is the cause of event *c*, so causation is a transitive relation.

Not all philosophers think, however, that these answers are correct. One argument that is advanced here is that, as Kurt Gödel showed **(Ref.)**, there are solutions to Einstein's equations of General Relativity in which spacetime forms a closed temporal loop. What this means is that there will be moments of time – call them t_1, t_2, and t_3 – such that while t_1 is earlier than t_2, and t_2 is earlier than t_3, t_3 is also earlier than t_2. Moments of time are, the, like points on a circle, and as one goes around the circle in one direction, one passes first through time t_1, then through time t_2, then through time t_3, but then one passes once again through time t_1. So time t_1 is earlier than t_3, but time t_3 is also earlier than time t_1. In addition, we have that time t_1 is earlier than t_1. Consequently the earlier than relation is neither asymmetric nor irreflexive.

Einstein's Theory of General Relativity is, however, logically compatible with a completely deterministic universe. So suppose that the laws of physics were such both that General Relativity was true, and that the total state of the universe at any moment causally determined how the universe would be at any later moment. Then the total state of the universe at time t_1 would be the cause of the total state of the universe at time t_3, but the total state of the universe at time t_3 would also be the cause of the total state of the universe at time t_1. Then, by the transitivity of causation, it would follow that the total state of the universe at time t_1 would be the cause of the total of the universe at time t_1. Hence, the relation of causation is neither asymmetric nor irreflexive.

Does this argument give one a good reason to believe that causal loops are logically possible, and thus that causation is not asymmetric? The answer is that it does not. For suppose that one showed both that a certain analysis of causation was correct, and that it followed from that analysis that causal loops are logically impossible. The conclusion would then be that the solution to the equations of General Relativity that Gödel discovered was, as it turns out, logically possible. What Gödel showed was not that a spacetime involving a temporal loop is compatible with the laws of nature if General Relativity is true, but, instead, the conditional conclusion that if a temporal loop is logically possible, then a spacetime involving such a loop is compatible with Einstein's Theory of General Relativity.

Suppose, on the other hand, that temporal loops are logically possible. Then there would, I think, be good reasons for concluding that causal loops are also logically possible. But that conclusion would still be compatible with the view that causation involves an asymmetric and irreflexive relation. For there could be a relation of direct causation, understood as a causal relation that satisfies the following constraint:

Necessarily, for all x and y, if x directly causes y, then there is no z such that x directly causes z and z directly causes y.

It could also be the case that, necessarily, if x is the direct cause of y, then y is not the direct cause of x. In that case, direct causation would be an asymmetric and

irreflexive relation. But provided that the relation of direct causation was not a transitive relation, one could still have causal loops involving that relation. The upshot is that, even if temporal loops are logically possible, it does not follow that there cannot be a causal relation that is asymmetric and irreflexive.

Let us turn, now, to the second of the above questions. How can one attempt to show that causation has certain formal properties? If the concept of causation is analytically basic, then there is no way that one can demonstrate, for example, that causation is an asymmetric relation. So if one holds both that causation is asymmetric – or irreflexive, or transitive – and that the concept of causation cannot be analyzed, one will have to view the fact that causation has the formal property in question as a truth known by intuition, rather than one that can be supported by argument.

Suppose, then, that the concept of causation is analyzable. Then the door is open to the possibility of proving that causal loops are logically impossible, and there have been three main routes that have been pursued. First, some analyses of causation – such as David Hume's and Patrick Suppes' **(Ref.)** – involve the notion of temporal priority, so that event *c* can be a cause of event *e* only if *c* is earlier than *e*. Then, if one can show that the earlier-than relation is both transitive and asymmetric, that will entail that there cannot be temporal loops, so if temporal priority is a necessary condition of causation, causal loops will also be impossible, and so one will have a proof that causation is an asymmetric and irreflexive relation.

A second possible approach involves the idea that if causal loops are logically possible, then what may be called undercutting causal loops must also be logically possible. Consider, then, a situation where an object *a* exists at time t_1, at location *s* and where that state of affairs causes it to be the case at a later time t_2 that a time machine sets of on a journey back to time t_2, and a location near *s*, at which point it explodes and destroys object *a*. If such a causal loop existed, a contradiction would be true, since it would be true both that *a* existed at time t_1 and that it did not exist at time t_1. So causal loops must be logically impossible, since if they were possible, undercutting causal loops would be possible as well, and the latter entail contradictions.

A slight variant on this is that of oscillating causal loops. Here things will proceed as in the previous case, with the addition that the situation is such that the non-existence of object *a* at time t_1 will bring it about that the time travel machine will be reprogrammed before setting off, at time t_2, on its journey into the past, so that now what it will do is to bring into existence, at time t_1 and in location *s*, an object that is indistinguishable from object *a* with respect to its properties. The result will be that as one, as to speak, goes around and around the causal loop, the proposition that there exists, at time t_1 and in location *s*, an object that is indistinguishable from object *a*, will oscillate endlessly between truth and falsity.

The third main type of approach is associated with analyses of causation in which probability plays a role. Sometimes, as with Mellor **(Ref.)**, the contention is that causes raise the probabilities of their effects; in other cases, such as Tooley **(Ref.)**, the claim is that the *a posteriori* probabilities of effects is fixed by the *a priori* probabilities of their causes. But in either case, the basic contention is that if causal loops are logically possible, then the probability of a given effect is never settled, either increasing indefinitely, or oscillating, as one goes endlessly around the causal loop.

2.4 Causation and Time

Another very important group of issues concerns what relations are possible, or necessary, between causal relations between states of affairs and the temporal order of those states of affairs. Of these, the most important issue here is whether it is logically possible for a cause to be later than its effect. If one thinks of a cause as bringing its effect into existence, it is very natural to think that, necessarily, a cause cannot be later than its effect, since how could an event *c* cause an event *e* at a time when event *c* does not itself exist? But, as we shall discuss later, this line of thought rests upon a controversial view concerning the metaphysics of time – the view, namely, that, on the one hand, the past is ontologically fixed, whereas the future is ontologically open.

In addition, many science fiction stories, such as those in the *Terminator* movies, involve people traveling back in time. So it would certainly seem that one can imagine what time travel into the past would be like, and if that is so, does not that provide grounds for concluding that time travel is logically possible? But if one travels back into the past, then one comes to have memories about the future, about the experiences that one had before traveling into the past. A memory of an event, however, must be caused by the event in question. So if time travel into the past is logically possible, so are causal relations running from later events to earlier events.

Another very important issue is this: Is it logically possible for a cause and its effect to be simultaneous? Here, too, if one thinks of a cause as bringing its effect into existence, it is very natural to think that it is impossible for a cause and its effect to be simultaneous. For while that will not be a case where one event causes another at a time when the first event does not yet exist, if one thinks of causation as one event bringing another into existence, it is natural to think that the cause must first exist, followed by the effect.

We shall see, however, that an interesting argument can be offered for the view that a cause can be simultaneous with its effect. Indeed, it has been argued not only that a cause and its effect *can* be simultaneous, but also that they *must* be **(Ref.)**.

Both the issue of whether a cause can be later than its effect, and the issue of whether a cause and its effect can be simultaneous are related to the question

of the formal properties of causation, and one way of attempting to show that neither of these things is possible is by arguing that both entail that causal loops are possible, but that causal loops are not logically possible.

Another, though less important issue, concerns whether temporally gappy causation is logically possible. Could an event *c* at time t_1 cause an event *e* at some different time t_2 without is being the case that, for every time *t* between t_1 and t_2 there is some event *d* such that *c*'s causing *e* is ontologically based upon *c*'s causing *d* and *d*'s causing *e*? Some philosophers, such as Wesley Salmon *(Ref.)* have offered analyses of causation according to which two temporally distinct events can be causally related only if they are connected by a continuous causal process. But is this requirement justified?

A final important issue involving time and causation concerns is this: Are the direction of time and the direction of causation logically independent or not? If not, is the direction of time to be defined in terms of the direction of causation, or does the direction of time enter into the analysis of the direction of causation?

As noted earlier, some philosophers **(Ref.)** have advanced analyses of causation according to which the direction of causation is based upon the direction of time. Other philosophers **(Ref.)** have offered analyses of the earlier-than relation according to which the direction of time is based upon the direction of causation.

2.5 Causation and Space

Finally, there are questions concerning relations between causation and space, and here there are two important questions. The first, which parallels a question about causation and time, is this: Is it logically possible for a cause and its effect to be spatially separated, without there being any causally intermediate state of affairs?

As noted earlier, some analyses of causation entail that causally related events in different locations must be connected by continuous causal processes, and this requirement excludes both temporally gappy causal processes and spatially gappy ones.

The other important question concerns whether spatiotemporal points or regions can themselves stand in causal relations. Why might one think that this is the case? One reason is this. Consider any two instantaneous slices of the total spatiotemporal world. What makes it the case that those two instantaneous slices belong to one and the same spatiotemporal world? One answer to this question is that they belong to one and the same world because they are causally connected.

Another reason for thinking that spatiotemporal points or regions can themselves stand in causal relations is this. First of all, as I shall argue, there is much that is appealing about the idea that the direction of time is based upon the

direction of causation. Secondly, however, there are also serious objections to such a view. But thirdly, those objections can be answered provided that there can be causal relations between spatiotemporal points or regions.

3. Alternative General Approaches to the Nature of Causation

As noted earlier, questions concerning the nature of causation first came sharply into focus as a result of David Hume's famous discussions (1739-40 and 1748). Since that time, many different accounts have been advanced. It will be helpful, however, to divide these various approaches up into four general types – - which I shall refer to as (1) direct, non-reductionist, (2) Humean reductionist, (3) non-Humean reductionist, and (4) indirect, or theoretical, non-reductionist – since, as we shall see, there are powerful general arguments that often bear upon all of the approaches within each of these general groups.

This fourfold division, in turn, rests upon the following three distinctions: first, that between reductionism and non-reductionism; secondly, that between Humean and non-Humean states of affairs; and, thirdly, that between states that are directly observable and those that are not. Let us consider, then, each of these distinctions in turn.

3.1 Non-Reductionism versus Reductionism

The non-reductionism-versus-reductionism distinction in this area arises in connection with both causal laws, and causal relations between states of affairs. As regards causal laws, reductionists claim that causal laws are reducible to facts about the total history of the universe, while non-reductionists deny that this is so. Similarly, as regards causal relations, reductionists claim that causal relations between states of affairs are reducible to non-causal facts about states of affairs, including the non-causal properties of, and relations between, events, whereas non-reductionists claim that no such reduction is possible.

But what exactly does reduction come to in these cases? The answer is that reductions can take two forms. On the one hand, there are analytical reductions, where the relations in question hold as a matter of logical necessity, broadly understood. On the other, there are reductions that involve a contingent identification of causation with some other relation.

3.1.1 Analytic Reductionism

A traditional way of formulating the basic issue in the case of analytical reductionism is in terms of whether the relevant causal concepts are *analyzable* in non-causal terms. It seems preferable, however, to formulate the relevant theses in terms of the slightly broader concept of *logical supervenience*.

The intuitive idea involved in the concept of logical supervenience can be explained as follows. Suppose that somewhere on our universe there was a

planet that was physically like our earth in all respects, and that, at every moment throughout its history had humans who, both physically and mentally, were exactly alike the humans living on earth at that time. In 1944 on earth, the United States was at war with Japan. Could it be that, on that other planet, the humans who, at that time, exactly corresponded to the humans on earth who were, respectively, citizens of the United States and citizens of Japan, did not belong to corresponding countries? Or could it have been the case that there were countries corresponding to the United States and Japan, but those countries were not at war at that time?

Neither of these things seems logically possible. But if that's right, then it seems that facts of a physical and psychological sort logically fix facts about nations. Any two worlds that agree with respect to all facts of the former sorts must, therefore, also agree with respect to facts of the latter sort.

In addition, the converse is not true. Our world could have been different in various ways, with regard both to what the earth was like physically, and the psychological states of humans, while all facts about nations were unchanged. So facts about nations do not logically fix all facts about the earth and its human inhabitants.

Moreover, there could be worlds with humans, but where there were no countries, since the humans had not entered into the relevant relations, whereas there could not be countries without bodies of land inhabited by humans or other types of persons.

A definition of logical supervenience can be modeled on this case. Let us say that facts of some type *S* (the supervening facts) *logically supervene* upon facts of some other type *B* (the supervenience base) if and only if the following things are the case:

(1) It is logically impossible for there to be any facts of type *S* unless there are at least some facts of type *B*.

(2) It is logically possible for there to be facts of type *B* even if there are no facts of type *S*.

(3) Let *X* be any total set of facts of type *S*. Then there at least two different total sets of facts of type *B* – call them *Y* and *Z* – such that *X* is logically compatible with *Y*, and also logically compatible with *Z*. So facts of type *S* do not logically determine what facts there are of type *B*.

(4) Let *X* be any total set of facts of type *B*. Then there exists only one total set of facts of type *S* – call it *Y*– such that *X* is logically compatible with *Y*. So facts of type *B* do logically determine what facts there are of type *S*.

Given the concept of logical supervenience, the relevant reductionist theses can be characterized as follows. In the case of causal relations between states of affairs, a thesis that is essential to reductionism is this:

Basic Reductionism with respect to Causal Relations

Any two worlds that agree both with respect to all of the non-causal properties of, and non-causal relations between, particulars, and with respect to all causal laws, must also agree with respect to all of the causal relations between states of affairs. Causal relations are, then, logically supervenient upon the totality of instances of non-causal properties and relations, together with causal laws.

But while this thesis is an essential part of a reductionist view of causation, it is not sufficient, since this thesis can be combined with a view of causal laws according to which they obtain in virtue of atomic, and therefore irreducible, facts. What is needed, then, is a reductionist thesis concerning causal laws, and here there are two important possibilities:

Strong Reductionism with respect to Causal Laws

Any two worlds that agree with respect to all of the non-causal properties of, and non-causal relations between, particulars, must also agree with respect to causal laws. Causal laws are, then, logically supervenient upon the totality of instances of non-causal properties and relations.

Moderate Reductionism with respect to Causal Laws

Any two worlds that agree both with respect to all of the non-causal properties of, and non-causal relations between, particulars, and also with respect to all laws of nature, must also agree with respect to all causal laws. Causal laws are, then, logically supervenient upon the totality of instances of non-causal properties and relations, together with laws of nature.

What lies behind this strong-reductionism-versus-moderate-reductionism distinction? The answer is that while most philosophers who are reductionists with regard to causation tend to identify laws of nature with certain cosmic regularities, it is possible to be a reductionist with regard to causation while holding that laws are more than certain cosmic regularities: one might hold, for example, that laws of nature are second-order relations between universals (Dretske, 1977; Tooley, 1977; Armstrong, 1983). Such a person would reject Strong Reductionism with regard to causal laws, while accepting Moderate Reductionism.

Each of these two reductionist theses concerning causal laws then entails, in conjunction with the Basic Reductionist thesis concerning causal relations, a corresponding thesis concerning causal relations between states of affairs:

Strong Reductionism with respect to Causal Relations

Any two worlds that agree with respect to all of the non-causal properties of, and non-causal relations between, particulars, must also agree with respect to all causal relations between states of affairs. Causal relations are, in short, logically supervenient upon the totality of instances of non-causal properties and relations.

Moderate Reductionism with respect to Causal Relations

Any two worlds that agree both with respect to all of the non-causal properties of, and non-causal relations between, particulars, and with respect to all laws of nature, must also agree with respect to all of the causal relations between states of affairs. Causal relations are, then, logically supervenient upon the totality of instances of non-causal properties and relations, together with laws of nature.

To be a reductionist with regard to causation, then, is to accept the Basic Reductionist thesis with respect to causal relations, and either the Strong or the Moderate Reductionist thesis with respect to causal laws. This then commits one either to the Strong Reductionist thesis or the Moderate Reductionist thesis with respect to causal relations.

3.1.2 Contingent Identity Theories

Analytical reductionism is not, however, the only reductionist possibility. For, just as in the case of philosophy of mind, where one can grant that mental states of affairs are not logically supervenient upon physical states of affairs, but then go on to claim that mental states of affairs are contingently identical with physical states of affairs, so one can reject analytical reductionism in the case of causation, but hold that causation in our world is, as matter of fact, identical with some relation that can be characterized in non-causal terms. Thus David Fair (1979), for example, has proposed that causation in the actual world can be identified with the transference of energy and/or momentum, while other writers, such as Wesley Salmon (1997) and Phil Dowe (2000a and 2000b), have suggested that causal processes can be identified with continuous processes in which quantities are conserved.

3.1.3 Non-reductionism

Finally, given the above accounts of analytical reductionism and contingent reductionism with respect to causation, non-reductionism with regard to causation can be defined as simply the rejection of those two alternatives. A non-reduction holds, accordingly, and first of all, that there is no combination of non-causal properties and non-causal relations with which the relation of causation is contingently identical. Secondly, with regard to analytical reductionism, the non-reductionist either rejects the Basic Reductionist thesis concerning causal relations, or else accepts that thesis, but rejects both the Strong and the Moderate Reductionist theses with regard to causal laws, or, finally, rejects all three of these analytical reductionist theses.

3.2 Humean Versus non-Humean Reductionism

In addition to the gulf between reductionism and non-reductionism, there are also very important divides within both reductionism and non-reductionism. In the case of reductionism, the crucial division involves a distinction between

what may be called Humean and non-Humean states of affairs. So let us now turn to that distinction.

One way of explaining the idea of a Humean state of affairs is contained, in effect, in the following passage from David Lewis's introduction to Volume II of his *Philosophical Papers* (1986, ix-x), where Lewis is explaining the idea of Humean supervenience:

"Humean supervenience is named in honor of the great denier of necessary connections. It is the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. (But it is no part of the thesis that these local matters are mental.) We have geometry: a system of external relations of spatiotemporal distance between points. Maybe points of spacetime itself, maybe point-sized bits of matter or aether or fields, maybe both. And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. For short: we have an arrangement of qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that."

Given this account of Humean supervenience, one can characterize Humean states of affairs as those that constitute the supervenience base. Humean states of affairs will then consist of states of affairs involving some spatiotemporal arrangement of point-sized entities and the natural, intrinsic properties of such entities.

This gives one a clear account, but it seems unsatisfactory for two reasons. First of all, it would not seem that those states of affairs exhaust the category that one is attempting to capture. Why so? One reason is that it would seem, first of all, that entities larger than a point can have intrinsic properties, that do not logically supervene on the intrinsic properties of, and spatiotemporal relations among, the point-sized parts of those larger regions, and, secondly, that this can be so without coming into conflict with Hume's denial of necessary connections between distinct existences.

One way of supporting this claim is as follows. Many people, both philosophers and non-philosophers, believe that complex states of the brain give rise, for example, to visual experiences, and that those visual experiences involve properties – such as qualitative redness – that are not reducible to the stuff of physics. If this is right, do such experiences have any location in our spatiotemporal world, and, if so, what is that location? I think that the most plausible view is that they do have spatiotemporal location, and that an experience is located where the complex neural state that causally gave rise to it is located. But the latter is not a point-sized region, and so the question arises as to whether the experience can, so to speak, be broken up into parts that are located at points within that region. This does not strike me as plausible, since it

does not seem to me likely that parts of the complex neural state causally give rise to parts of the experience.

Suppose, then, that experiences cannot be broken up into parts that stand in relations to point-sized parts of neural states. Then experiences as a whole – or at least non-point-sized parts of experiences – have to be viewed as causally related, in a non-reducible way, to non-point-sized parts of neural states. But would this entail the existence of any necessary connections between distinct existences? I cannot see how it would, and so it seems to me that Lewis's account of Humean supervenience is too narrow.

The second problem with Lewis's account is that it does not exclude states of affairs involving properties that entail necessary connections between distinct states of affairs. A point-mass, for example, could have the power of being indestructible. The crucial question is then whether that power can be identified with an intrinsic property of the point-mass. Many philosophers hold that it cannot be, since they hold that powers must be reduced to a combination of intrinsic properties plus laws of nature, and the latter are not intrinsic to an object. But some philosophers, as we shall see in chapter **(Ref.)**, hold that there can be irreducible powers that are intrinsic properties of objects. If this is possible, then a point-mass's have such an intrinsic property of being indestructible would entail the later existence of that point-mass, and so there would be a state of affairs that was Humean, on Lewis's account, but that entailed logical connections between temporally distinct states of affairs.

How, then, is the distinction between Humean and non-Humean states of affairs to be explained? Lewis's approach, in effect, was to attempt to define Humean states of affairs. I think that one should, instead, begin at the other end – that is, by defining non-Humean states of affairs.

But how is the latter to be defined? Consider the property of being indestructible. How can the idea of that property be defined if the property is an irreducible one? Presumably as follows:

P is the intrinsic property of indestructibility = def.

Property *P* is an intrinsic property such that, for any thing, *x*, *x*'s having property *P* at any time *t* logically entails that for every later time *t**, *x* exists at time *t**.

The crucial feature of this definition is that it involves the idea of *logical entailment*. For suppose that one attempts to express the idea without using that notion – for example, as follows:

P is the intrinsic property of indestructibility = def.

Property *P* is an intrinsic property such that, for any thing, *x*, if *x* has property *P* at any time *t* then, for every later time *t**, *x* exists at time *t**.

This latter definition does not say that if something has property *P*, then it is indestructible. What is says is rather that nothing that has property *P* ever, as a

matter of fact, ceases to exist. But this can be true without such a thing being such that it is impossible to destroy it.

My proposal, then, is that we generalize upon this feature of the definition of the concept of an intrinsic property of indestructibility to define what it is for the concept of a property or relation to be the concept of a non-Humean:

C is the concept of a non-Humean property or relation = def.

C is the concept of a descriptive, non-logical property or relation, and the analysis of concept *C* requires the concept of logical entailment.

A non-Humean state of affairs can then be defined as a state of affairs that involves one or more non-Humeans properties and/or relations, while a Humean state of affairs can then be defined as any state of affairs that is not non-Humean. A reductionist analysis is then a Humean reduction if all of the properties and relations involved in the reduction are Humean. Otherwise it is non-Humean.

Ultimately, when we turn to a consideration of approaches to causation that involve what are called 'propensities', we shall see that we need to expand the idea of non-Humean properties and relations. But that is best left for later.

Are there any non-Humean properties or relations? Hume did not employ that concept, but he held that there could not be logical connections between distinct existences. But if there were non-Humean properties or relations, then there would be such connections. So, though he did use the concept, we can say that Hume denied the existence of non-Humean properties and relations, and did so because he held that they are logically impossible.

Was Hume right about this? This is one of the most important questions in metaphysics, and especially with regard to causation and laws of nature. Many present-day philosophers share Hume's view that non-Humean properties and relations are logically impossible. If they are right, then unless the concept of causation is analytically basic, it must be susceptible of a Humean reductionist analysis. But if, as other philosophers hold, non-Humean properties and relations are logically possible, then very different approaches to the analysis of causation are open, both reductionist and non-reductionist.

3.3 Direct Versus Indirect Awareness

The third and final distinction is that between direct non-reductionism and indirect, or theoretical non-reductionism. According to direct non-reductionism, some causal states of affairs can be immediately given. Are these causal laws, or causal relations between states of affairs? Since it is not at all plausible that one can be directly acquainted with causal laws, the relevant states of affairs must consist of causal relations between states of affairs. Thus direct non-reductionism can be defined as a version of non-reductionism that claims that the relation of causation is immediately given in experience.

Indirect, or theoretical non-reductionism rejects this claim, maintaining either that the relation of causation is itself an irreducible, theoretical relation, or, alternatively, that causal laws are irreducible, theoretical states of affairs, and that causal relations must be reduced to causal laws, plus non-causal properties and relations. Either way, then, the relation of causation is not directly observable.

4. Two Crucial Underlying Issues

4.1 Causation and Analysis

1. Does the concept of causation stand in need of analysis?
2. If it does, what types of analysis are allowable?

4.2 Causation, Physics, and Metaphysics