6 Causation

- Hume's account of causation
- The response to Hume
- Neo-Humean approaches

Overview

Traditional metaphysicians took causation to be a modal notion; they held that causes necessitate their effects. Hume attacked this idea. Invoking an empiricist theory of concepts, he claimed that if the concept of causation did involve the idea of necessary connection, the necessity would be an empirically manifest feature of particular causal sequences, and he argued that it is not. Causation, he insisted, is just constant conjunction or regularity of succession. Defenders of the traditional approach respond to Hume in a number of ways. Some (like Kant) reject Hume's empiricism and insist that causation is an *apriori* concept. Others claim that Hume's argument establishes only that causation is not an observational notion; they hold that causation is a theoretical concept. Still others insist that the causal relation is one that can be directly observed. More typical, however, are those philosophers who endorse Hume's insistence that we provide a nonmodal account of causation. Among recent metaphysicians, some (like J. L. Mackie) continue to believe that a regularity analysis provides the requisite nonmodal account; whereas others follow David Lewis in defending a counterfactual analysis of causation.

Hume's account of causation

The concept of causation is about as central as any to our thinking about the world. We typically suppose that it is events that play the lead role in causal phenomena. Indeed, we think of causation as a relation between events: one event, we say, causes another; the first is cause; the second, effect.[1] The relation, we think, is a kind of glue that holds our world together, relating phenomena that would otherwise be separate and independent. It is also, we think, a kind of engine that keeps our world going: without the causal relation, there would be none of the changes or processes that make up the history of the world. And we think it plays these roles not just with regard to the physical world. Causation is a relation that spans the physical and the mental. Just as the tossing of a baseball causes the window to shatter, so my belief that it is raining and my desire to stay dry seem to provide a causal explanation of my taking my umbrella out of the closet before I walk to work.

And the notion of causation plays these roles both in our everyday thinking about the world and in the more specialized thinking at work in contexts like medicine, the law, and the various sciences. Both the centrality and pervasiveness of the concept make it a natural target for metaphysical analysis, and throughout the history of metaphysics it has been just such a target.

Traditional metaphysicians offer us quite different accounts of causation; but one theme that recurs in their accounts is the idea that what marks out an event as a cause is a special power, force, or energy. In virtue of that power, force, or energy, an event brings about another event – its effect; and it does so of necessity. The connection between cause and effect is, then, a modal connection. A cause necessitates its effect; it makes it happen. Given the occurrence of the cause, the event that is its effect must occur; it cannot fail to occur.[2]

So traditional metaphysicians tell us that a cause and its effect are tied together by a modal relation, a kind of necessary connection. But, then, given the pervasiveness of the causal relation, this traditional account of causation results in a picture of the world as shot through with modality. For the philosopher who is suspicious of modal notions, the traditional account is certain to appear problematic. As we mentioned in the last chapter, David Hume is a philosopher with deep suspicions about the idea that the world has genuinely modal features. Not surprisingly, he launched a major assault on the traditional account of causality.[3]

Central to Hume's attack on the idea of necessary connection is a certain claim about our ideas. The claim is that every idea has its origin in experience. Hume calls the immediate deliverances of experience impressions; and he tells us that there are two kinds of impressions. There are the impressions of sensation, impressions that result from turning our attention outwards to the objects making up the so-called external world; and there are the impressions of reflection, impressions that result from directing our attention inward to the introspective data of consciousness. So Hume thinks that all our ideas are traceable to the impressions of sensation and reflection. Indeed, he thinks that all of our ideas are either copies of impressions of one of these two sorts or composites made up of ideas that are copies of sensory or reflective impressions.

Now, Hume takes a dim view of the claims of traditional metaphysicians. Many of those claims, he thinks, are either unclear or straightforwardly unintelligible. The difficulty is that metaphysicians use language with no discernible empirical content; and he thinks that a paradigmatic example is traditional metaphysical discourse about causation. The central claim here, he thinks, is just the one we have laid out – that causes exert a special force or energy in virtue of which their occurrence necessarily brings about the occurrence of their effects. Hume, of course, objects to this claim, and the way he formulates his objection is by saying that there is no idea at all corresponding to the traditional metaphysicians' use of the phrase 'necessary connection.' Were we to have such an idea, he claims, it would be one for which we could

identify an empirical origin; that is, it would be an idea for which there is some corresponding impression of sensation or reflection. So to cast doubt on the very intelligibility of the traditional account of causation, Hume invites us to examine the individual cases of causation we meet in sensation and introspection; and what he argues is that our examination of those cases reveals no modal features – no power, no energy, no force, and no necessary connection.

He begins with the case of sensation, where we are confronted with bodies interacting with each other. The sort of case he has in mind is the familiar one where we have a first billiard ball striking a second and causing it to move. He claims that if we examine just a single case of this sort of interaction, we find that we have two events exhibiting a temporal relation. The event we call the cause – the first ball's striking the second – precedes the event we call the effect – the second ball's moving. We observe, then, a temporal succession in the events. Furthermore, we observe that the two balls are in contact at the moment the first strikes the second. So we have what Hume calls impressions of temporal succession and spatial contiguity; but Hume insists that these are the only relations we experience when we examine the interaction. In particular, we do not observe anything corresponding to the traditional metaphysician's talk of power, energy, or necessitation. We see one event succeeding another in a narrowly circumscribed region of space; and that is all we see. Were we to have a sensory impression of some special causal power on the part of the first event, then on our very first acquaintance with an instance of this sort of sequence, our experience of the first event in the sequence would enable us to infer just which event would follow. And, of course, our observation of any such novel sequence tells us no such thing; nor, Hume tells us, is this an accident. Where we have, as we do in our example, two separate and distinct things, the items are completely independent; there is no inference ticket from the one to the other. That, Hume wants to claim, is just what their being distinct and separate consists in.

But if sensation provides us with no impression of any causal power or force necessitating an event, perhaps the introspective case does. Here, we meet the phenomenon of volition; and attending to that phenomenon, it can seem, provides us with a first person awareness of the sort of power or energy traditional metaphysicians speak of. I am sitting in my chair. I have been dozing, but I remember that I need to wash the dishes. Accordingly, I decide to get up and go to the kitchen. I am, however, sleepy; and it is difficult to rouse myself, so I focus on the intended action. I exert my will, and my body responds. Is this not a case where I have direct and immediate access to the causal power or efficacy that results in the necessitation of an event?

Hume thinks not. He tells us that what we have in the interaction just described is an experience of a mental event followed by a physical event and nothing more. I could not have first hand knowledge of a tie or connection between the two events without understanding how the mental and the physical interact. Hume, however, reminds us that the relation between mind

and body is utterly mysterious and that no philosopher has done the first thing to dispel the mystery. Furthermore, he tells us, if I did directly apprehend the necessitating connection tying my act of the will with my body's rising from the chair, I would have first hand knowledge of every intermediate item in the chain of events taking us from the volition to the movement of the body; but, of course, not even the most sophisticated physiologist knows what all those items are. More to the point, from the perspective of the first person phenomenology of volition, none of us apprehends any of the events intermediate between an act of the will and the body's responding. What we observe is simply a succession of events. In the introspective case, then, we meet even less than in the case of the billiard balls. In the latter case, we have both temporal succession and spatial contiguity. Since we do not grasp our mental events as things having a spatial location, in the volition case we experience only the temporal succession between cause and effect. And in neither case, do we have an impression of any kind of energy, force, power, or necessary connection. But, then, since every idea derives from an impression of experience, we seem driven to the conclusion that we have no clear and coherent idea of a necessitating connection between a cause and its effect; and that suggests that the traditional metaphysician's talk of causal force, causal power, and the like is deeply confused if not completely meaningless.

But however it may be with traditional metaphysicians, the suggestion that the ideas of temporal succession and, perhaps, spatial contiguity exhaust our idea of causation is bound to appear problematic. Certainly, events can bear those relations to each other without being related as cause and effect. Hume agrees; he thinks that there is something more to causation; and he thinks that to discover the missing ingredient, we need to enlarge our field of observation. We need to look beyond our sample causal sequence to cases where we have events resembling the cause in our sample. We look, for example, beyond the case of the two billiard balls to other cases where a moving object strikes an object of roughly the same size and mass. What we find, in each case, Hume tells us, is that the second object moves. So we find that events resembling our original cause are associated with events resembling our original effect. Furthermore, we observe that, in each case, the two events are related in precisely the way our original cause and effect were: we have the relevant temporal succession and the relevant spatial contiguity; and, of course, we find ourselves labeling the temporally prior event cause and the event succeeding it effect. But in none of these cases do we find anything that was missing in our original case. What, then, is it that makes all of these sequences causal?

Hume's answer is that while taken individually none of the sequences exhibits any feature that might justify calling them causal, the sequences all conform to a general pattern. We have two sets of resembling events; and in each sequence, an event from the one set bears the relevant temporal and spatial relations to an event from the other set. More precisely, there are two kinds of events, K_1 and K_2; and in each sequence an event belonging to K_1 is

succeeded by a spatially contiguous event belonging to K_2, so that we can say that whenever an event from K_1 occurs, a spatially proximate event from K_2 will follow. And according to Hume, this is all that there is to causation. Causation is nothing more than the sort of constant conjunction at work in the pattern. Accordingly, when we say that one event causes another, we are not pointing to any feature of the events that, taken in isolation, they can be observed to exhibit. We are saying, instead, that our events instantiate a general pattern of the sort just identified.

So our idea of causation involves no modal notions. Causation is simply constant conjunction or regularity of succession. But Hume thinks that when we appreciate this fact, we are in a position to see why traditional metaphysicians mistakenly thought that necessitation is a component in our idea of the causal relation. Once experience has provided us with evidence of a causal pattern, a pattern in which events of one sort are regularly followed by spatially proximate events of another sort, the observation of an occurrence of an event of the first sort creates an expectation of an event of the second sort. Indeed, for anyone familiar with the pattern, the mere thought of an event of the first sort leads to a thought of an event of the second sort. In both cases, the mind is, as Hume puts it, directed from the cause to the effect; and that direction, he thinks, is what leads traditional metaphysicians to talk of power, force, energy, and necessitation. They are, of course, confused: they are construing a purely subjective feature of our thinking about causal sequences as an objective feature of those sequences themselves. Why do they make that mistake? In a very famous passage, Hume gives his answer:

> The mind has a great propensity to spread itself on external objects and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time that these objects make their appearance to the senses.[4]

So it is because we human beings tend to project our subjective reactions to phenomena onto the phenomena themselves that we think that causes necessitate their effects. Now, this propensity of the mind to be carried from an impression or idea of a cause to an idea of its effect may strike us as an interesting, but in the end, accidental feature of the causal relation. It is significant, however, that Hume himself wants to build the propensity into the definition of causation or, at least, into one definition of causation. He rounds off his discussion of causation by offering two different definitions of the notion of a cause. The first involves merely the idea of constant conjunction:

> We may define a cause to be 'an object precedent to and contiguous to another and where all the objects resembling the first are placed in like relations of precedence and contiguity to those objects that resemble the latter'.[5]

Then, he tells us that if we wish, we may substitute for this definition the following:

> A cause is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other.[6]

The response to Hume

So Hume thinks that where we have a causal sequence, we have an instance of a pattern of constant conjunction between events of two kinds, a pattern that determines the mind to move from an experience or idea of an event of the one kind to the idea of an event of the other kind. Causation is, then, a thoroughly nonmodal relation. Those who disagree have raised a variety of objections to Hume's analysis. One kind of objection is that the account is too broad. Thus, critics point to noncausal patterns in which events of one kind are regularly followed by events of another kind. Thomas Reid gives the day-night sequence as just such a pattern.[7] The arrival of night invariably follows the termination of day, and yet we refuse to say that day causes night. One reason is that we could just as well say that night causes day. That claim, however, would deliver the consequence that one and the same event is related to another event as both cause and effect, and we believe that cause and effect are asymmetrically related; we believe, that is, if an event, c, is the cause of an event, e, then e is not a cause of c. Another such example is outlined by A. C. Ewing.[8] A horn goes off at a certain factory in London at 8:00 a.m. each day; immediately after, the workers at another factory in Manchester enter their factory and begin work. The horn, of course, is meant to signal the start of work at the London factory and not the Manchester factory; but Ewing argues that if the regularity account were true, we would be forced to say that the sounding of the horn in London is no less the cause of the workers entering the Manchester factory than it is the cause of the corresponding event at the London factory.

A second kind of difficulty for the regularity theory is presented by singular causal judgments. If Hume were right, then all such judgments are implicitly general in the sense that when we say that some individual episode, c, causes some individual episode, e, the truth of our claim presupposes the possibility of identifying some kinds, K_1 and K_2, such that, first, c and e are members, respectively, of K_1 and K_2 and, second, instances of K_1 are invariably followed by instances of K_2. But, of course, for many perfectly appropriate causal claims, it is preposterous to suppose that there are any such general claims lying in the background. Historical claims are one obvious case. The claim that the assassination of Archduke Ferdinand caused the First World War is a true causal claim, but hardly one that unfolds into any plausible Humean generalization.

Furthermore, if Hume were right about causation and regularity, we could

never be justified in arriving at a causal judgment on the basis of a single experience of a succession of events. We would, on the contrary, need to experience a large number of instances. Critics, however, point out that we are often able to make a causal judgment on the basis of just a single instance of succession. Thus, suppose I present you with a weird looking contraption of a sort you have never seen before, and suppose that upon my striking it, bells, whistles, and lights from within the contraption all go off. You will certainly be justified in asserting that my striking the contraption caused the reaction.[9]

Philosophers who are sympathetic with the traditional metaphysician's idea that causation involves concepts like power, force, energy, or necessary connection typically find these sorts of objections telling; but, of course, they need to reply to Hume's attack on the sort of approach they endorse. The central objection is that since we have no experience of the power, force, energy, or necessary connection that is supposed to characterize the causal relation, we must reject the traditional metaphysician's account. Anti-Humeans respond to the objection in a number of ways. One response is to concede that the idea has no empirical origin, but to reject the consequence Hume draws from this fact. One might argue that a strongly modal conception of causation is presupposed by anything we might call experience and conclude that such a notion is an *apriori* concept, that is, a concept that is not derived from experience. Such, at least, is Kant's response to Hume.[10] Kant holds that a presupposition of our having the sort of unified or coherent experience we do in fact have is that disparate events are related in some rule-governed way that makes it possible for us to infer the occurrence of one event from that of another. Accordingly, no event can be an object of experience for us unless it stands in a strongly modal causal relation to other events. Accepting Hume's argument that no such modal notion can have an empirical origin, Kant concludes that the concept of causation is innate. It is one of twelve *apriori* concepts or categories that understanding imposes on the raw data of inner and other senses (sensation and introspection) to yield what we call an object of experience.

We find sympathy with the Kantian view that causation is a strongly modal relation that cannot have an empirical origin among early twentieth-century idealists. They attempt to develop the view by providing a substantive characterization of the causal relation. They tell us that it is a relation which is somehow analogous to the logical relation of entailment.[11] Just as the premises of a valid argument necessitate their conclusion, so a cause necessitates its effect. That the causal relation is analogous to the logical relation, we are told, is shown, first, by the fact that we can infer effects from their causes and, second, by the fact that causes provide us with reasons or explanations for their effects. Both facts are intelligible only on the assumption that causation is something like entailment.

So one response to Hume is to insist on a modal characterization of causation and to construe the relation as an *apriori* or innate concept. To endorse this sort of response is to reject the thesis that empiricism provides an

adequate account of the derivation of all the ideas that can figure in human cognition. A somewhat different objection to Hume does not necessarily require a general rejection of empiricism, but merely a rejection of the very austere form of empiricism that Hume presents. According to this objection, what Hume established is merely that a modal notion of causation is not an observation concept; that is, that it is not a concept whose application gets warranted exclusively by reference to sense experience or introspection. Hume challenges us to identify something in experience (whether inner or outer) that we can intelligibly construe as the relation of necessary connection or causal efficacy, but proponents of this second response deny that our failure to meet this challenge calls into question the idea that causality is at bottom modal. They tell us that if we generalize Hume's challenge, we get the result that virtually none of our theoretical concepts has any sort of legitimacy. Concepts like that of an electron, a quark, a muon and a gluon all fail Hume's test: nonetheless, our best physics makes essential use of these concepts, and physics represents the paradigm of a successful intellectual enterprise.

Theoretical concepts, then, do not submit to the model of concept formation that lies at the core of Hume's extreme form of empiricism; and, according to this second reply, causation is a theoretical concept.[12] We never directly experience the causal relation; it is rather a relation we postulate. Like all theoretical concepts, the notion has its origin in a whole battery of interrelated conceptual moves including extrapolation, analogy, and inference to the best explanation; and like other theoretical notions, it gets justified by the explanatory work it does, and all of these conceptual moves can be accommodated by a less austere, more enlightened form of empiricism. According to proponents of this second reply, what the postulation of the causal relation explains is, among other things, precisely the phenomenon Hume mistakenly identified with causation. On this view, Hume was not wrong to associate regular succession with causation. Typically at least, causal relations between events give rise to regular sequences. Regularity of sequence is not, however, causation, but rather a symptom of the existence of a strongly modal relation of causation. That relation, we are told, is not observable, but it issues in and serves to explain regular sequences that are observable.

A final reply defenders of the traditional account have made is to insist that a strongly modal notion of causation counts as an observation concept. The proponent of this reply will claim that we can literally observe causal efficacy, that we can directly experience one thing's making another happen. Some who endorse this claim maintain that Hume is simply wrong about the phenomenology associated with volition. They maintain that we are introspectively aware of causal efficacy when we undergo acts of volition that culminate in action;[13] whereas others insist that we can literally perceive causal efficacy in the world around us.[14] These theorists point to the experience of ordinary cases where one thing pushes, pulls, or strikes another thing. They tell us that these are all cases where one thing or event generates, produces, gives rise to another, and the claim is that they represent the

paradigmatic cases where one thing makes another happen. On this view, it was Hume's attachment to an impoverished model for understanding sense perception that led him to deny that we can be perceptually aware of causal efficacy. He took sense experience to have as its objects things like colors, sounds, smells, and shapes; but the claim is that if we restrict the range of sense experience to things like these, we will find ourselves denying not simply that we can experience necessary connections. We will be forced to deny that we can perceive the very things – billiard balls, logs, rocks – that Hume tells us enter into the regular sequences he calls causal. To accommodate our experience of the world, we need a broader notion of perception, one that allows us to say that we perceive not just familiar concrete particulars, but the physical changes, processes, events, and interactions into which they enter as well.

Neo-Humean approaches

But while there are philosophers who want to defend a broadly modal account of causation, they have been in the minority. The more popular stance among recent metaphysicians is one of sympathy with Hume's overarching aims. Like Hume, these philosophers reject the idea that causation is an irreducibly modal relation and seek an alternative to the traditional account. Some of these philosophers further agree with Hume that the best hope for a nonmodal account is a regularity analysis; but they think that we need to supplement Hume's analysis to fortify it against counterexamples; or they think that we need to provide a completely different formulation of the insight that causation is to be understood in terms of regularity. Others who are sympathetic with Hume's aims think that we need to scuttle the regularity approach if we are to provide a satisfactory nonmodal account of causation. We need to look at examples of both kinds of nonmodal analysis.

Defenders of the regularity approach owe us a reply to the various objections raised at the beginning of the last section. There were, recall, two types of objections: one bearing on regular, but noncausal sequences, the other, on singular causal judgments. Defenders of the Humean regularity approach typically find the second type of objection less serious. They insist that where we make a causal judgment on the basis of what appears to be a single experience, we are not making a singular judgment about a genuinely novel case. Our judgment involves an assimilation of the case before us to some familiar pattern where we already have the requisite Humean regularity. Thus, in the case of the contraption, we have familiar cases where manipulation is followed by some observable result – my flipping the switch and the light going on, my pushing the button and the television screen being illuminated, my pulling the lever and the door opening. The singular judgment that my striking the contraption caused the sound and light show is simply an expression of the belief that the relevant sequence is an instance of the familiar pattern.

The other kind of difficulty – that bearing on noncausal, yet regular successions – has played a more central role in post-Humean attempts at defending some sort of regularity analysis of causation. John Stuart Mill thought that Reid's example of day and night shows that Hume's account needs to be supplemented.[15] Besides being invariable, Mill says, a genuinely causal regularity is unconditional; that is, it holds no matter what. It does not hinge on conditions that need not obtain. The day/night case, however, fails this test; or so Mill says. He tells us that the sequence would fail if the sun were to be extinguished or if the earth were to cease rotating in the appropriate way, and he concludes that it obtains only conditionally.

A different strategy for dealing with regular, but noncausal sequences was proposed by the Logical Positivists of the 1920s, 1930s, and 1940s. As they saw it, the really serious threat to a regularity analysis is the case of merely accidental sequences of the sort at work in the case of the two factories, and they claimed that what distinguishes a genuinely causal succession of events from a merely accidental correlation is that the former has the status of a law of nature or is derivable from something that has that status.[16] For this approach to succeed, of course, the regularity theorist needs to come up with an account of the notion of a law that does not involve the concept of causation. The Positivist tradition is replete with attempts at the requisite kind of analysis. Some of these attempts focused on the logical form of sentences expressing laws (the so-called lawlike sentences), stressing the unique syntactical or semantical properties of such sentences. Others stressed the pragmatic role that laws play in the explanation and prediction central to the ongoing activity of the overall scientific enterprise.

Another approach to causation that has been popular with regularity theorists invokes the notions of necessary condition and sufficient condition. Among these accounts, certainly the most influential is that of J. L. Mackie.[17] Mackie is concerned with the causal claims we actually make, whether in specialized contexts like the sciences and medicine or in the nonspecialized context of everyday life. He thinks that such claims always presuppose a background setting – what Mackie calls a causal field. The causal field represents the context in which we take our cause to operate; it is the region within which the cause makes a difference. According to Mackie, causal claims are responses to causal questions, and those questions are typically incomplete and indeterminate. Giving those questions a complete and determinate content is a matter of identifying a causal field. When we ask, for example, why this or that individual contracted cancer, we may be asking why the individual contracted the disease now rather than at some earlier time; in that case, the causal field is the lifetime of the individual. We may, however, be asking why this individual contracted cancer when other individuals who were also exposed to the asbestos in the factory did not; in that case, our causal field is those human beings who were exposed to the asbestos in a particular factory.

So a causal claim is always issued relative to a particular causal field. But

what are we saying when we issue a claim circumscribed in this way? Mackie presents us with the example of a house fire. The experts examine the house after the fire has been extinguished, and they tell us that the cause of the fire was an electrical short-circuit. According to Mackie, they are not telling us that the short-circuit was a necessary condition of the house fire; they know that any of a large number of other factors could have resulted in the house catching fire at the time it did; nor are they saying that the short-circuit was a sufficient condition for the fire. They know that the short-circuit by itself was not sufficient to set the house afire. A lot of other factors had to be in place: the dry rags had to be there next to the electrical outlet; the water sprinklers had to be defective; and so on.

So what the experts pick out as the cause of the fire is neither a necessary nor a sufficient condition for the fire. It is rather an indispensable component in a larger bundle of factors, all of which were present and which, taken together, were sufficient for the fire. There are, of course, other such bundles of factors sufficient for producing the same result; but none of them was present before the fire. What the experts are calling the cause, then, is an *insufficient*, but *necessary* component in a bundle of factors that was *unnecessary*, but *sufficient* for the occurrence of the fire. Mackie calls such a factor an *INUS* condition, where the term is built out of the initial letter in each of the terms italicized above; and Mackie wants to claim that what we typically call a cause is just the sort of thing the experts are calling the cause of the fire – an INUS condition. Thus, to identify the cause of an event relative to a given causal field is to specify some factor that is an insufficient, but necessary component in one of the bundles that, within that field, are not necessary, but are sufficient for the occurrence of the event, to say that all the other factors in that same bundle were present, and to deny that any of the other bundles sufficient for the event were present.

As we indicated, this account is supposed to be a version of the regularity approach to causation. But it will qualify as a regularity account only if we can understand the account's talk of necessary and sufficient conditions in regularity terms. Towards showing us that we can, Mackie proposes that we understand his talk of necessary and sufficient conditions in terms of certain conditional statements. Thus, he proposes that we understand the claim

(1) Event x was a necessary condition for event y

in terms of the counterfactual conditional

(2) If x had not occurred, y would not have occurred;

and he recommends that we understand the claim

(3) Event x was a sufficient condition for event y

in terms of what he (following Nelson Goodman) calls the factual conditional

(4) Since *x* occurred, *y* occurred.

As attempts to display the nonmodal character of talk about necessary and sufficient conditions, however, these proposals can strike us as disappointing. Not only do they not seem to display the INUS condition account as a version of the regularity approach; they suggest that at bottom the account is a modalist theory. After all, do we not need modal notions to make sense of (2) and (4)? Mackie thinks not. He thinks that we can understand the conditionals he points to as condensed or telescoped argument forms. Thus, (2) is to be understood as

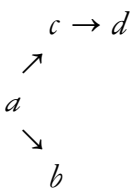
(5) Suppose *x* did not occur; then *y* did not occur;

and (4) as

(6) *x* occurred; therefore *y* occurred.

Of course, these arguments need to be fleshed out. They both need additional premises; but Mackie assures us that those premises will turn out to be straightforwardly Humean generalizations, nonmodal regularity statements; and Mackie insists that a speaker can make a claim of the form of (1) or (3) without being able to specify precisely which generalizations are required to complete the telescoped argument that underlies the claim.

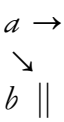
So we have regularity approaches to causation; but as we indicated earlier, not all nonmodal analyses of causation involve the regularity strategy. Indeed, the most influential recent account of causation is that of David Lewis, who recommends that we understand the phenomenon in counterfactual terms.[18] Lewis has doubts about the prospects for a successful regularity approach. He mentions a number of problems for the approach; and although he does not mention Mackie by name, Lewis seems to have the INUS condition account in mind when he sets out the problems. One of the problems bears on what Lewis calls epiphenomenal effects. Suppose that an event, *a*, causes two different events, *b* and *c*. Suppose further that *b* has no causal consequences; *b* is a causal deadend or an epiphenomenal effect of *a*. Suppose finally that the other effect of *a* – *c* – causes some further effect, *d*.



Since it is in virtue of the laws and circumstances that *a* causes both *b* and *c*, *b*

turns out to be an INUS condition of *d*; for given the laws and the circumstances, *b* is an insufficient, but necessary component in a bundle of factors (*a*, *b*, *c*) that is sufficient but, we assume, not necessary for the occurrence of *d*. Accordingly, despite the fact that *b* is a causal deadend – an event with no effects whatsoever, *b* turns out to be a cause of *d* on Mackie's INUS condition account.

Lewis mentions another problem for an account like Mackie's. This is the problem of causal pre-emption. Here, we have two events, *a* and *b*, each of which, taken by itself, would cause a third event, *c*. However, when *a* and *b* both occur, *a* acts to block *b*'s normal causal role and goes on to cause *c* all by itself.



So *b* occurs, but is pre-empted by *a*. However, since the circumstances are such that had *a* not blocked *b*, *b* would have caused *c*, *b* is an insufficient, but necessary component in a bundle of factors (including *b* and the relevant circumstances) that while sufficient for *c* is not necessary for *c*. So *b* is an INUS condition for *c*; but it is not the cause of *c*, so, again, a counterexample to Mackie's analysis.

In the face of these difficulties, Lewis recommends that we approach the analysis of causation by way of the notion of counterfactuality. Although he takes the counterfactual analysis to be a rival to regularity accounts, Lewis thinks, first, that it is a properly nonmodal account and, second, that it is a genuinely Humean approach. We have already noted that counterfactual discourse appears to be a form of modal discourse; but in the last chapter we saw how Lewis invokes a nominalistic and thoroughly nonmodal account of possible worlds as concrete particulars in providing a reductive analysis of a wide range of modal phenomena. He wants to claim that the same strategy can be employed in providing a reductive or nonmodal account of counterfactual conditionals. And Lewis thinks that the idea that causation is to be understood in counterfactual terms is one we meet in Hume himself. In the first section of this chapter, we quoted two different definitions of causation that Hume presents in his early work, the *Treatise on Human Nature*. In a later work, the *Enquiry into Human Understanding*, Hume once again rounds off his discussion of causation with two definitions, one that identifies causation with constant conjunction and the other that makes reference to the movement of the mind from cause to effect; but the first of these definitions adds something that we do not find in the corresponding definition in the *Treatise*. Hume says:

> we may define a cause to be an object followed by another and where all the objects similar to the first are followed by objects similar to the second. *Or*

> *in other words where, if the first object had not been, the second had never existed*.[19] (italics mine)

Lewis points to the sentence I have italicized; and he remarks that while Hume takes the sentence to be an alternative formulation of what we meet in the first sentence, the sentence actually summarizes a completely different form of analysis – a counterfactual as opposed to a regularity analysis of causation.

Now, like Mackie, Lewis is concerned with a broad use of the term 'cause' of the sort at work in both specialized contexts and everyday life. In that use, what we call the cause of an event is just one of a number of different factors relevant to the event's occurrence. Our identification of one among these factors as the cause hinges on our interests and purposes in inquiry. Given those interests and purposes, we call the other factors mere conditions; but with a different set of interests and a different causal field in mind, one or more of the other factors relevant to the occurrence of our event could have been singled out as cause.

So it is a broad notion of cause that Lewis tries to capture in his counterfactual analysis. In its most general form, what the analysis is telling us is just what Hume tells us in the concluding sentence in the passage just quoted from the *Enquiry*: to say that a certain event, c, causes another event, e, is to say that if c had not occurred, e would not have occurred. But, of course, the trick for anyone anxious to give a nonmodal account of causation is to do what Hume does not do – to show that we can give a properly nonmodal account of the sort of counterfactual at work in this claim. Now, in general, Lewis wants to claim that a counterfactual conditional issued in a world, w, is a claim about what goes on at another possible world, a world that while different from w, is like w in important ways. Accordingly, towards giving us the required nonmodal analysis of counterfactuals, Lewis introduces a notion of comparative similarity among possible worlds. The idea is that one world, w_1, can resemble another world, w_2, more than some third world, w_3, does. Lewis tells us that the factors relevant to judgments of comparative similarity include the particular matters of fact that obtain in the various worlds as well as the laws of nature that hold at those worlds; nevertheless, he refuses to provide a formal definition of the notion, telling us that he takes the relation to be primitive.

As I have said, Lewis takes a counterfactual conditional issued in a world, w, to be a claim about how things go at a world that bears certain similarity relations to w; but since it is counterfactuals issued in our world – the actual world – that will be relevant to the causal claims we actually make, Lewis invites us to focus on an ordering of worlds according to their comparative similarity to our world. In this ordering, we move from worlds that are less like ours to worlds that are progressively more like ours. Given this ordering, we have the resources for giving an account of just when a counterfactual of the form

(7) If it were the case that *p*, then it would be the case that *q*

is true. We begin by singling out the *p*-worlds (that is, the possible worlds where the proposition that *p* is true) and the *q*-worlds (that is, the possible worlds where the proposition that *q* is true). Then, if we endorse the assumption that there is such a thing as the *p*-world that is closest to or most similar to our world, we can say that (7) is true just in case the *p*-world closest to our world is a *q*-world; that is, just in case it is true that among all the possible worlds where *p* is true, the one that most resembles our world is a world where *q* is true.

Now, some defenders of a possible worlds theory of counterfactuals endorse this assumption and so accept the analysis just formulated.[20] Lewis, however, is suspicious of the assumption. He thinks it hazardous to assume that for any proposition, *r*, there is an *r*-world that is closest to or most resembles our world. It might be, he thinks, that for any world, *w*, there is a world, *w'* that resembles our world more than *w* does. Accordingly, he tells us that a counterfactual of the form of (7) is true just in case there is a *p*-world, *w*, such that *q* is true in *w* and *w* resembles our world more than any *p*-world where *q* is false.

Our concern, of course, is with causal relations between events, and we have seen that the core idea behind the counterfactual approach to causation is that where an event, *c*, causes an event, *e*, that fact is to be understood in terms of the counterfactual conditional

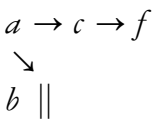
(8) If *c* were not to occur, *e* would not occur;

but when we apply Lewis's account of the truth conditions for counterfactuals to (8), we get the result that (8) is true just in case there is a possible world, *w*, such that neither *c* nor *e* occurs in *w* and *w* is closer to the actual world than any possible world where *c* does not occur, but *e* does. When a proposition like (8) is true, we can say that *e* causally depends on *c*. Now, there can be chains of events linked by this relation of causal dependence. Thus, we might have a chain of events *a*, *b*, *c*, *d* . . . , where *b* is causally dependent on *a*, *c* on *b*, *d* on *c*, and so on. Lewis calls such a chain a causal chain and tells us that one event, *c*, causes another event, *e*, just in case there is a causal chain leading from *c* to *e*. So causation is to be understood in terms of causal dependence; causal dependence is to be understood in terms of counterfactuals; and counterfactuals are to be understood in terms of the ordering of possible worlds imposed by the relation of comparative similarity. However, since as Lewis understands them, possible worlds are just concrete particulars – things that can be understood in straightforwardly nonmodal terms, the account is, in the intended way, thoroughly nonmodal.

But do not the problems Lewis sets out for the regularity approach infect his own counterfactual analysis? Consider the problem of epiphenomenal effects. An event, *a*, causes two events, *b* and *c*; *c*, in turn, causes some fourth

event, *d*; whereas, *b* is an epiphenomenal effect – an effect that has no effects. Now, suppose that given the laws and the circumstances, *d* could not have come about except by way of *a* and *c*. But, then, it seems that, given the laws and the circumstances, if *b* had not occurred, its cause, *a*, would not have occurred either; and that means that neither *c* nor *d* would have occurred. Accordingly, we get in result that if *b* had not occurred, *d* would not have occurred, so that, our causally inert or merely epiphenomenal effect seems to turn out, once again, to be the cause of *d*.

Likewise, certain kinds of cases of causal pre-emption seem to be a problem for Lewis. Suppose that events *a* and *b* both occur. Taken in isolation, each of *a* and *b* would cause an event, *f*, each by way of an intermediate cause, with *a* causing *f* by way of *c* and *b* causing *f* by way of *d*. But, when both *a* and *b* occur, *a* acts to block the occurrence of *d*, so that *f* gets caused by way of the causal chain *a*, *c*, *f*.



But, then, it should turn out, on Lewis's analysis, that if *c* had not occurred, its effect – *f* – would not have occurred either. That, however, seems to be false since if *c* had not occurred, neither would its cause, *a*, have occurred; but, then, there would have been no causal pre-emption and *b*'s effect, *d*, would have occurred and caused *f*. The result seems to be that, on Lewis's account, we cannot say that *c* causes *f*, when, of course, it does.

Lewis, however, insists that neither case represents a genuine counterexample to his account. In both scenarios, we get an untoward result only because we assume that, in causal contexts, it is legitimate to invoke what Lewis calls backtracking counterfactuals. Backtracking counterfactuals are conditionals that make what happened in the past counterfactually dependent on what happens at a later time. Thus, in first the case, we assumed that if the epiphenomenal effect, *b*, had not occurred, its cause, *a*, would not have occurred. Likewise, in the pre-emption case, we assumed that if the intermediate cause, *c*, had not occurred, its cause, *a*, would have not occurred. Lewis, however, rejects both claims because he thinks that it is illegitimate to employ backtracking counterfactuals. Their use presupposes that the past is dependent on the present and future. Lewis denies that this is the case.[21] How things went in the past is not counterfactually dependent on how they will go in the future. On the contrary, the future is dependent on the past and present. But if we reject the backtracking assumptions, neither the case of the epiphenomenal effect nor the case of causal pre-emption constitutes a problem for Lewis's account.

There remains, however, one kind of case where critics have argued that Lewis's account does not deliver the results that we might want. The case is that of causal overdetermination, the case where two potential causes operate

simultaneously to produce an effect that either would have produced without the other. Thus, two individuals simultaneously shoot bullets into a man's heart and the man dies. On Lewis's account, the man's death is not counterfactually dependent on either bullet, and we get the result that neither is the cause of the man's death. Lewis, however, defends himself by denying that cases of overdetermination should be used as test cases for a theory of causation. He thinks that they represent cases where our intuitions give out. We just do not know what to say about these cases, so it cannot represent a flaw in a theory of causation that it fails to give us an unambiguous verdict on cases of causal overdetermination.[22]

Notes

1 Sometimes, however, we seem to speak of things from other categories as items that play the causal role. Thus, we sometimes seem to be saying that substances are causes. This happens most often in connection with rational agents or persons, and it has led some metaphysicians to develop theories of agent causation. See Chisholm (1964), Taylor (1966), and O'Connor (2000). The vast majority of metaphysicians, however, take causation to be a relation between events. In this chapter, I focus on their work.
2 See, for example, Aristotle, *Metaphysics* IX.5 (1048^a 5–7).
3 The texts which provide the focus for my discussion are Hume (1739: book I, part III, section XIV) and Hume (1748: section VII).
4 Hume (1739: 167).
5 Hume (1739: 172).
6 *Ibid*.
7 Reid's attack on Hume's account is found in Reid (1788: essay 4).
8 Ewing (1951: chap. VIII). One might argue that Ewing's example is not a genuine counterexample to Hume's analysis on the grounds that the sounding of the horn and the workers entering the factory in Manchester lack the requisite spatial proximity. Ewing, I suspect, would respond by pointing out that the case of mental causation (where we have an event that has no spatial location at all) shows that spatial proximity of cause and effect is not an essential feature of causation. Accordingly, Ewing would say, all that remains of Hume's analysis is regularity of temporal succession.
9 The example is taken from Ducasse (1951: 91–100).
10 See Kant (1787: Second Analogy, 218–33).
11 See, for example, Ewing (1951: chap. VIII).
12 For a discussion of this approach, see Tooley (2003: 425–30).
13 See, for example, Armstrong (1997: 319–28).
14 See, for example, Anscombe (1971). Anscombe denies that causation involves necessitation. She thinks that effects derive from or arise out of their causes, and she thinks that we can perceive the derivation of the effect from its cause.
15 For Mill's theory of causation, see Mill (1843: vol. I, book 3, chaps 4–6 and vol. II, book 3, chap. 21.
16 See, for example, Schlick (1932).
17 The most detailed presentation of this account is found in Mackie (1965).

18 For Lewis's account, see Lewis (1973). Further elaboration of the view is found in Lewis (1986b).
19 Hume (1748: 51).
20 See, for example, Stalnaker (1968).
21 Obviously, much more needs to be said in defense of the prohibition against backtracking counterfactuals. See Lewis (1979), which is reprinted in Lewis (1986a: vol. II), together with a postscript.
22 A rather different objection against Lewis's account is that since there are counterfactual conditionals that have nothing to do with causal determination, counterfactuality is too broad a notion for an analysis of causation. See Kim (1973) for a statement of this objection.

Further reading

Hume's discussions of causation in Hume (1739) and Hume (1748) are essential reading for anyone interested in the metaphysics of causation. For anti-Humean approaches, I would recommend the discussions of causation in the Second Analogy of Kant (1787), chapter VIII of Ewing (1951), and Anscombe (1971). Mackie (1965) presents the most influential regularity account of causation to be found in recent literature; and Lewis (1973) provides a clear, if not altogether easy presentation of the counterfactual approach. Finally, the April 2000 issue of the *Journal of Philosophy* brings together papers exhibiting very recent thinking on causation.