18 Causation, Counterfactuals, and the Third Factor

Tim Maudlin

In his classic paper "Causation" (1973a), David Lewis urged that regularity accounts of causation, having continually failed to resolve the counterexamples and problematic cases with which they were faced, be abandoned in favor of an alternative strategy: the analysis of causation in terms of counterfactuals. More than a quarter of a century later, the time has come to evaluate the health and prospects of Lewis's program in its turn.

No one would now deny that there is some deep conceptual connection between causation and counterfactuals. As Lewis points out, Hume himself (incorrectly) paraphrased one of his analyses of causation using a counterfactual locution (Hume 1902 [1748], sec. VII):

> ... or in other words, *where if the first had not been, the second had never existed.*

In honor of Hume, let us call the counterfactual "If C had not occurred, E would not have occurred," when both C and E did actually occur, the *Hume counterfactual*. When we think we know a cause of some event, we typically assent to the corresponding Hume counterfactual. Furthermore, our interest in causes often has a practical aspect: We want to know the causes of events so that we can either prevent or foster similar sorts of events at other times. Causal claims are therefore deeply implicated with the sorts of future subjunctives used in practical deliberation: If we should do X (which we might or might not, for all we now know) then the result would be Y. The future subjunctive is a close cousin to the counterfactual conditional, since accepting a future subjunctive commits one to accepting the corresponding Hume counterfactual in the event that the antecedent does not come about.

But Hume's dictum hardly survives a few minutes contemplation before counterexamples crowd into view. The sort of counterfactual dependency Hume cites is not necessary for causation: Perhaps the effect still would have occurred despite the absence of the cause since *another* cause would have stepped in to bring it about. The dependency is also not uncontroversially sufficient for causation. If John Kennedy had not been assassinated on November 22, 1963, he would have still been president in December 1963. But surely too, if Kennedy had still been president in December 1963, he would not have been assassinated in November of that year: When asked to consider the latter counterfactual we do not imagine Kennedy killed in November and resurrected in December. So the counterfactual dependencies go both ways and the causal arrow only one.

These sorts of problems have been thoroughly chewed over in the literature, and no detailed review of the strategies used to respond to them is needed here. The problem of backup (potential) causes can perhaps be solved by using counterfactual dependency to define direct causation and then taking the ancestral to define causation, or by making the effect *qua* effect very fragile in some way and arguing that the backup could not have produced it at exactly the right time or right place or with all of the right properties, or by appealing to a matching strategy, in which the cause that is operative when the backup is present is identified by matching it to an otherwise identical case where the backup is absent (and the Hume counterfactual holds). Attempts to fix the counterfactual analysis have become ever more subtle, and complicated, and convoluted (as perhaps some other essays in this volume will attest). I have neither the space not the desire to address them all individually.

I do not want to quarrel with these sorts of solutions in detail because I want to argue that the attempt to analyze causation in terms of counterfactuals of this sort is wrongheaded in a way that no amount of fine-tuning can fix. Causation is not to be analyzed in terms of counterfactual dependency at all, no matter how many equants and epicycles are appended to the original rough draft.[1]

If causation is not to be analyzed in terms of counterfactual dependency, how are we to explain the systematic connections between judgments about causes and judgments about counterfactuals? Such connections may be secured analytically if causation can be defined in terms of counterfactuals, but I am denying that this can be done. Connections would also obtain if counterfactuals could be reduced analytically to causal claims, but that is an even less appetizing project than the one we are rejecting. The only other possibility is that a third factor is involved, some *component* of the truth conditions of counterfactual claims that is also a component of the truth conditions of causal claims. This third factor would provide the analogue of a "common cause" explanation for the systematic connections between causal claims and counterfactuals: Neither underpins the other, but the third factor underpins them both.

The prospects for a "third factor" explanation obviously depend on the identification of the missing factor. I think it can be identified: What links causation and counterfactuals by figuring in the truth conditions for both is natural law. Laws play one role in determining which counterfactuals are true, and another role in securing causal connections. The "necessary connexion" that Hume sought at the heart of causation is nomic necessity.[2]

1 Knowledge of Causation without Knowledge of any Hume Counterfactual

Let us consider what one must know, in at least one case, in order to know what caused an event. Suppose you know that the laws that govern a world are the laws of

Newtonian mechanics. And suppose you also know that forces in this world are all extremely short range: forces exist only between particles that come within an angstrom of each other. And suppose particle *P* is at rest (in an inertial frame) at t_0 and moving at t_1, and that in the period between t_0 and t_1 only one particle, particle *Q*, came within an angstrom of *P*. Then, I claim, we know with complete certainty what caused *P* to start moving: It was the collision with *Q*.

Thus: Given that we know the laws and we know some very circumscribed set of particular facts, we know the cause of *P*'s motion. But do we know what would have happened if *Q* had not collided with *P* (i.e., if *Q* had not approached within one angstrom of *P*)? We do not. Suppose, for example, that the vicinity of our particles is chock-a-block full of monitoring equipment, which tracks the exact trajectory of *Q*, and jammed with particle-launching devices loaded with particles just like *Q* and designed to launch these particles so as to collide with *P* just as *Q* did if *Q* should deviate in any way from its path. There are (for all I have told you) innumerable such devices, monitoring the path at arbitrarily close intervals, all prepared to step in should *Q* fail to collide with *P*. I hereby warn you of the presence of these contraptions, unspecified in number and construction, and ask you now whether we know what would have happened if *Q* had not collided with *P*.

We do not have enough information to evaluate this counterfactual, both because the *way Q* fails to collide has not been specified and because the exact construction and disposition of the monitoring devices has not been indicated. Perhaps for many sorts of deviation, some other particle, just like *Q*, would have collided with *P* at just the same time and in just the same way. Perhaps this is even true for *all* possible deviations, or all the sorts of deviation we would consider relevant. So perhaps we could be convinced that *P* would have moved at exactly the same time and in just the same way even if *Q* had not collided with it. Still, *none of this has anything at all to do with the fact that the collision with Q is what caused P to move*. The existence of the monitoring devices and potential backup particles is simply *irrelevant* to the claim that the collision with *Q* was the cause. In fact, once we know the laws we don't even *care* what would have happened if *Q* had not collided with *P*: Perhaps *P* would not have moved, or perhaps it would have because something else collided with it. The information we have (viz., the laws of nature in this world) allows us to identify the cause without knowing that *any* Hume counterfactual is true.

The counterfactual analyst can respond in several ways. One would be to insist that we know the causes in this case because we know that *some* Hume counterfactual is true, although (owing to the lack of information) we don't know which one. No matter how many backup systems there are, there must be some point on the trajectory of *Q* such that had it miraculously swerved at that point, *P* would not have moved since none of the backups would have had a chance to fire. But since I have not specified the number or location of the backups, how does one know this is true? What if there is an infinite sequence of backups, progressively faster and faster reacting, monitoring at ever closer and closer intervals?

Lewis (1986b) addresses this sort of criticism, claiming that we need not pay much heed to our intuitions about such recondite circumstances. Discussing exactly this sort of problem (as raised by William Goosens) and a case involving action at a distance, he responds: "I do not worry about either of these far-fetched cases. They both go against what we take to be the ways of this world; they violate the presuppositions of our habits of thought; it would be no surprise if our common-sense judgments about them went astray—spoils to the victor!" (1986b, p. 203). Lewis's strategy here deserves some comment. The rules of the game in this sort of analytic project are relatively clear: Any proposed analysis is tested against particular cases, usually imaginary, for which we have strong intuitions. The accuracy with which the judgments of the analysis match the deliverances of intuition then constitutes a measure of the adequacy of the analysis. Unfortunately, it is often the case that the question of *how* the intuitions are arrived at is left to the side: Getting the analysis to deliver up the right results, by hook or by crook, is all that matters, and this, in turn, encourages ever more baroque constructions. But if we care about intuitions at all, we ought to care about the underlying mechanism that generates them, and in this case it seems plausible that a simple *argument* accounts for our judgment, *an argument entirely unaffected by the existence of backup devices, no matter how numerous*. The case does differ from those of our actual experience, and in such a way that the envisioned situation will surely never occur. But the *way* that the example differs from more familiar ones *makes no difference whatsoever to the reasoning that allows us to identify the cause*: So long as none of the backup devices *fires*, their number and complexity is perfectly irrelevant. We have such strong intuitions about the case even though it is far-fetched because the respect in which it is far-fetched makes no difference to the method by which the cause is identified. The example shows that causes are not identified via the Hume counterfactual.

What one would like to say, of course, is that the following counterfactual it true: If Q had not collided with *P and none of the back-up devices had fired*, then P would not have moved. And this is correct. So we want to hold fixed, in evaluating the counterfactual, *the nonfiring of potential alternative causes*. But in order to know what to hold fixed, we already have to know quite a lot about the (actual and counterfactual) causal structure, since we have to identify the possible-but-not-actual-alternative-causes to hold fixed. But now the project looks pretty hopeless as a way of using counterfactuals to get a handle on causation: We already have to bring in causal judgments when determining which counterfactual to consider.

So if we do not know that the collision with Q caused P to move by knowing what would have happened if Q had not collided with P, how do we know it? The argument is quite simple and straightforward, but it is worth laying out. Since the laws of this world are by hypothesis those of Newtonian physics, we know that particle P, which is at rest at t_0, will remain at rest unless some force is put on it. And since the forces in this world are all short range, we know that no force will be put on P unless some particle approaches within one angstrom of it. And since P does start to move, we know that some particle did approach within one angstrom of it, which is why it started moving. And since Q was the *only* particle to approach closely enough, we know that it was the collision with Q that caused P to move. End of story.

There are some counterfactuals implicit in this reasoning. If *nothing* had put a force on P it would not have accelerated. That, of course, follows from Newton's first law of motion. And if we knew that if Q had not put a force on P *nothing else* would have, then we would have the Hume counterfactual. But we do not know this. So as it stands we can only infer: If Q had not collided with P, either P would not have started moving or something else would have collided with it. This is not the Hume counterfactual; hence the headaches for the counterfactual analysis.

Various sorts of sophistication of the original simple counterfactual analysis might succeed in getting the case right, once enough detail is given. For example, if Q had not collided with P, but some other particle had instead, then it does seem plausible that P would not have moved off in quite the same way as it did. This seems plausible even though we cannot begin to specify *how* P's motion would have differed. It also seems plausible that there is *some* moment in Q's trajectory late enough that, had Q miraculously disappeared, it would be too late for a backup to fire—even though we have no clue about *how late* that point must be. So maybe these sorts of analysis can get the right result. But the curious thing is that even if they do get the right result, *it is obviously not on the basis of that result that we judge Q to be the cause. Rather, it is because we already judge Q to be the cause that we think that, with enough detail, we would get the right result*. This is obvious since our judgment that Q is the cause is secure *even before the details are filled in*. And the advantage of the little argument rehearsed above is that it explains how we know that Q was the cause, given the laws of motion and the fact that only Q collided with P, without having to know any more details about the situation. So even if it does turn out that had Q not collided, P would have at least moved differently, that is irrelevant (given the laws) to identifying the cause.

Since it is facts about the laws that help us identify the cause in this case, and since laws are obviously deeply implicated in the evaluation of counterfactuals, I suggest that we stop trying to analyze causation directly in terms of counterfactuals and consider anew how laws play a role in determining causes. At one end of the spectrum we have seen how the laws of Newtonian mechanics can enable one to identify the cause in the example discussed above, even if one does not know any Hume counterfactual. Let us now look at an example from the other end of the spectrum.

2 Knowledge of Counterfactuals without Knowledge of Causes

If an analysis of causation in terms of counterfactuals were correct, then, unless the analysis itself contains some vague concepts, fixing determinate truth values for all counterfactuals should fix the truth values of all causal claims. Of course, in many circumstances counterfactuals themselves may not have classical truth values: The antecedent or consequent of the conditional may be too vague. But supposing that the truth values of all relevant counterfactuals are sharp, then one would suppose that an analysis of causation in terms of counterfactuals should render the causes sharp. But we can imagine situations in which all relevant counterfactuals are determinate but causal claims are not, because it is unclear exactly what the laws are. The example is rather artificial, but illustrative.

Recall the rules of John Conway's Game of Life. Life is played on a square grid, using discrete moments of time. At any moment, each square in the grid is either empty or occupied. At any given moment of time, whether a square is occupied or not depends on the how that square and the eight immediately adjacent squares were occupied at the previous moment. For example, if four or more of the adjacent squares are occupied at one instant then the given square will be empty at the next instant, and if one or none of the adjacent squares are occupied then at the next moment the square will be empty. Conway's rules cover all possibilities, so the state of the grid evolves deterministically through time.

Now imagine a world in which space and time are discrete, and in which space consists in a rectangular grid of points that persist through time. And imagine that there is only one sort of matter in this world, so that every discrete part of space is, at any moment, either empty or occupied by the matter. And imagine that the patterns of occupation evolve in definite deterministic patterns similar to those in Conway's game, but somewhat less orderly. As in Conway's game, one can predict with certainty whether a given point will be occupied or not at an instant if one knows only the previous pattern of occupation of that point and the eight adjacent points. But unlike Conway's game, the rules of evolution cannot be distilled into a few simple rules.

In principle, the rules of evolution must specify for each of the 512 possible patterns of occupation of a 3-by-3 grid whether the central square is or is not occupied at the next instant.[3] Imagine picking the rules in a random way: For each of the 512 possible input patterns, flip a coin. If the coin lands heads, the central square will be occupied the next instant, if tails it will be empty. One now has a complete set of rules that will determine how patterns evolve through time. Now imagine a physical world of infinite extent and with a completely variegated state of occupation whose evolution in time always conforms to these rules.

(It is a key point that we are now imagining a world whose evolution everywhere *conforms to* these rules, but we are reserving judgment about whether we ought to say that the evolution is *generated by* these rules.)

If we have chosen the rules randomly, then it is overwhelmingly likely that there will be patterns of occupation that differ only in a single location, but which both yield the same result. For example, suppose it turns out that both patterns *A* and *B* in figure 18.1 are always succeeded by the central square being occupied. The question I want to ask is this: In such a world, is the bit of matter in the bottom central location of pattern *A* a *cause* (or an essential part of a cause) of the central square being occupied the next instant or not?

In the course of this inquiry, I will take a few points to be uncontroversial. First, I take it that the sort of physical world I have described at least *could be* a world that follows laws of the evolution of the distribution of matter. On anything like the Mill–Ramsey–Lewis account of laws, I take it that if the world is extensive and variegated

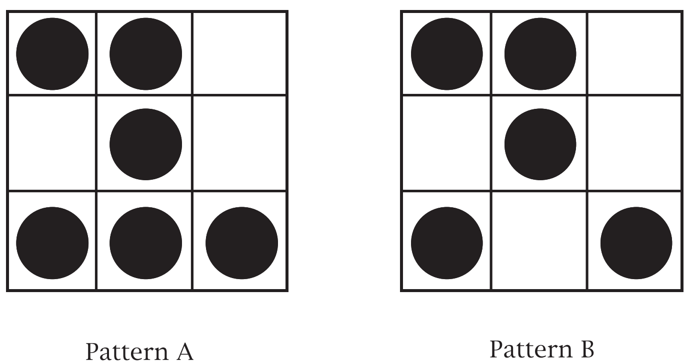

Figure 18.1

enough, then there *must be* laws: After all, of all possible distributions of matter throughout space and time, only a set of measure zero will conform to these rules through all time. It is therefore extremely informative to know that these rules are always obeyed. I further assume that having granted that the world is law governed, we grant that the truth values of all counterfactuals concerning the evolution of distributions of matter are determined. Since the rules are deterministic, we can use them to determine how any distribution of matter would have evolved, even if that distribution never actually occurs in the world we are imagining. So at the fundamental physical level, there is no dispute about counterfactuals: Given any complete pattern at a time, we know how it would have evolved had it been different in some precisely specified way.

My central claim is this: Even though there is no dispute, at the fundamental level, about counterfactuals, there can still be a dispute about causation, and further, this dispute about causation arises as a consequence of a parallel dispute about laws. The issue here is not how the case ought to be judged, or what our intuitions are about the case, or whether a particular theory gets the intuitions right. The issue is rather that unproblematic knowledge of all counterfactuals in this case does not seem to settle the question of causal structure, and furthermore, unproblematic knowledge of counterfactuals does not settle what the laws are in this same case. This again suggests that it is the laws, not the counterfactuals per se, that underwrite causal claims.

There are two lines of argument that can be addressed to the question of whether the bit of matter in the bottom central location of pattern A is a cause of the succeeding occupation or not. The first line makes straightforward appeal to the Hume counterfactual, and denies that the matter in that location is a cause. It is beyond dispute in this case that the Hume counterfactual is false: Even if the bottom central location had not been occupied (i.e., even if what was pattern A had been instead pattern B), still the central square in the succeeding instant would have been occupied. In this sense, it is beyond dispute that the presence or absence of matter in that location makes no difference to how the distribution of matter will evolve. Furthermore (this argument continues) although sometimes the Hume counterfactual can fail to be true for genuine causes and effects, that is only because causation is the ancestral of *direct* causation, and in direct causation the Hume counterfactual is always true. But since space and time are discrete here, and there are no intervening moments of time or intervening places in space, if there is any causation it must be direct causation. There is simply no place for there to be a chain of direct causes between pattern A and the succeeding occupation of the central square. So since there is no indirect causation (no room for it) and there is no direct causation (the Hume counterfactual is false) there is no causation at all linking the matter in the bottom

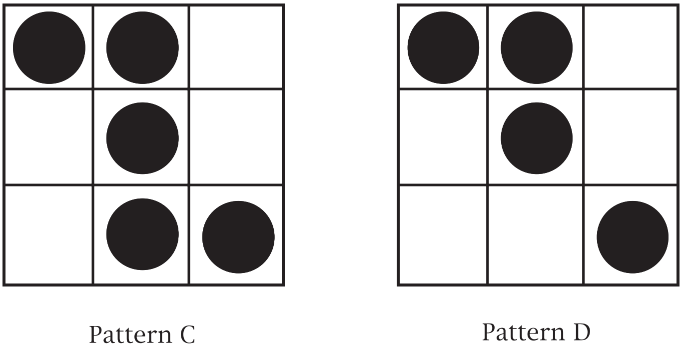

Figure 18.2

central location to the matter occupying the central square in the next instant. Hence the case against causation.

Now the case for causation. In arguing in favor of the matter in the bottom center being a cause, one would first point out that in general the laws of evolution in this world seem to link complete 3-by-3 patterns of occupation to the succeeding state of the central square. If the state of the central bottom location were *never* relevant to the evolution, then there would be little dispute about its causal efficaciousness. But suppose that patterns *C* and *D* in figure 18.2 are such that *C* is always followed by an occupied central square and *D* by an empty one. In this case, the Hume counterfactual holds, and the matter in the bottom central square of pattern *C* would be among the causes of the occupation of the central square in the next instant. The idea is that the complete set of 512 transition rules is so miscellaneous that there is no significantly *more* compact way to convey it than by specifying each of the 512 rules separately. It does not, for example, make things significantly simpler to try to bundle patterns *A* and *B* together into *one* rule (which refers to only the pattern of occupation of the eight locations other than the bottom central one) since one would have to indicate, either explicitly or implicitly, that in this case the state of the bottom central location need not be specified, although in other cases it must be.

So the advocate of causation argues as follows. The overall structure of the evolution in this world suggests that the laws of nature connect complete 3-by-3 patterns to the succeeding state of the central square. (After all, that is how *we* generated the rules: by flipping a coin anew for each of the 512 patterns.) Since there is no interesting generic pattern to the rules, the most *uniform* thing to do is to regard the fundamental laws of this world as all rules that take complete 3-by-3 patterns as input and give either an occupied or empty central square as output. From this point of view, it is just a *coincidence* that pattern A and pattern B, which differ only in the state of the bottom central square, both lead to occupied central squares. (After all, if we generate the rules by flipping a coin, it *is* just a coincidence that they do.) As far as the laws of nature are concerned, pattern A and pattern B have *nothing in common at all*; they are, as it were, each atomic and distinct. From this point of view, then, the state of the central bottom square is of vital importance, for it determines whether the transition falls under the law governing pattern A or the law governing pattern B, where these *laws* are regarded as fundamentally distinct. If an A pattern gives rise to an occupied central square, then although it is true that the square would still have been occupied had the bottom central location been empty, that would have been a case of an *alternative and distinct cause of the same effect*. The B pattern is a sort of backup mechanism, on this view, which would produce the same effect as the A pattern had the A pattern been different in a specified way. But whether the transition from the B pattern counts as *alternative mechanism* or as an instance of the *same* mechanism as the transition from the A pattern depends on what the laws are that govern the situation. Both transitions could be instances of the operation of the same law, or instances of two distinct laws, and our causal judgments depend on which we take to be the case.

We saw above that one route to save the counterfactual analysis was to add something to the antecedent: that all potential backup causes be held fixed as non-firing if in fact they did not fire. We saw that this spells trouble for trying to use the counterfactuals to analyze causation. Now we are seeing that what counts as an alternative or backup cause may depend on the nature of the laws governing the situation, so it is not puzzling that analyses that leave the laws aside will run into trouble.

I do not wish to adjudicate between the arguments pro and con the causal effectiveness of the bottom central location in this case; I rather simply want to acknowledge that there can be a dispute, a *reasonable* dispute, in a case like this where the counterfactuals are all beyond question. I do think, as a matter of fact, that the pro causation argument would eventually win out, if no interesting generic metarules could be found that simplified the presentation of the 512 different transitions. But even if one thinks that the con side would win, it is hard to deny that each side *has a point*, and that the point ultimately turns on how one is to conceive of the laws of nature in such a world.

We can easily imagine the 512 different transitions being such that they can be encapsulated by very simple rules, such as the Conway rules mentioned above. And we can imagine intermediate cases, where instead of giving all 512 transitions separately, one can give a few very complicated, generic rules. And we can imagine the case discussed above, where no interesting simple metarules exist. There is a slippery slope between these cases. I take it that on Lewis's account of laws, there will be points along this slope where it will be indeterminate what the laws are: points where there will be reasonable disagreement about whether the fewer complicated generic rules are *simpler overall* than the 512 separate transition rules. And where the laws are in dispute, it may turn out (as it does above) that the *causes* are in dispute, all while the truth values of the counterfactuals remain unquestioned.

So our pair of examples displays a nice symmetry. In the Newtonian case, we showed that we can know the causes without knowing any Hume counterfactuals if only we know the laws. And in our modified Game of Life, we have shown that we can know all the counterfactuals without being sure about the causes if there is a dispute about the laws. Between the two of them, they make a strong case for the idea that it is the laws rather than the counterfactuals that determine the causes.

3 The Role of Laws

If we are to explain the ubiquitous connections between causal claims and counterfactuals by appeal to the role of laws in the truth conditions of each, we must at least sketch how those truth conditions go. In the case of counterfactuals the basic story is extremely obvious and requires little comment. If I postulate that a world is governed by the laws of Newtonian mechanics, or the laws of the Game of Life, then certain conditionals are easy to evaluate: those that postulate a particular complete physical state at a given time as the antecedent and some other particular or generic physical state at a later time as the consequent. Given the state specified in the antecedent, the laws then generate all later states, and the sort of state specified by the consequent either occurs or does not. That is, one uses models of the laws as possible worlds and then employs the usual possible worlds semantics for the truth conditions of the conditional.

If the antecedent is not stated directly in the physical vocabulary, or is vague or generic, or is only a partial specification of the state of the world, then various sorts of filling in must be employed to yield a set of states that satisfy the antecedents, which then generate a set of models, and again things proceed in the usual way. There is much to be amplified in this sketch, and emendations needed for probabilistic laws, and so on, but the role of laws in the whole process is evident enough. Let us omit any more detail about counterfactuals, then, and turn to the much more interesting topic of causation.

I do not think that there is any *uniform* way that laws enter into the truth conditions for causal claims, as there is a fairly uniform way they enter into the truth conditions for counterfactuals. Rather, I think that laws of a very particular form, wonderfully illustrated by Newton's laws of motion, support a certain method of evaluating of causal claims, whereas a more generic and somewhat less intuitive use must be found for other sorts of laws. As we will see, the laws of the Life world lack all of the interesting characteristics of Newton's laws, so we will again use them for illustrative purposes.

Let's start, then, by returning to our colliding particles. Why are we so confident in identifying the collision with particle Q as the cause of particle P being set in motion? Recall the exact form of Newton's laws, the form that makes them so useful as aids for tracking causes in a case like this. The first law, the *law of inertia*, states that a body at rest will remain at rest and a body in motion will continue in motion at a uniform speed in a straight line, unless some force is put on it. The first law specifies *inertial motion*, that is, how the motion of an object will progress if nothing acts on it. The second law then specifies how the state of motion of an object will *change* if a force is put on it: It will change in the direction of the force, and proportionally to the force, and inversely proportionally to the mass of the object. Note that the second law is in a way parasitic on the first: The first specifies what is to count as a state of motion (uniform motion in a straight line), and the second how, and in what circumstances, the state changes.

The structure of Newton's laws is particularly suited to identifying causes. There is a sense, I think, in which the continuation of inertial motion in a Newtonian universe *is not caused*. If a body is at rest at one time, and nothing acts on it (i.e., no force acts on it), then it sounds odd to ask what causes it to remain at rest. It sounds odd to say that the body's own inertial mass causes it to remain at rest, since there is no force that the mass is resisting, and the inertial mass is just a measure of a body's resistance to force. And it sounds odd to say that the law of inertia itself causes the body to remain at rest, since it seems similar to a category mistake to ascribe causality to those very laws. Of course, the body remains at rest because no force acts and because the inertial state of motion of a body at rest is to remain at rest, but it certainly sounds odd to cite the absence of forces as a cause of remaining at rest.

Or at least, if there is any cause of a body at rest remaining at rest in a Newtonian universe it is a sort of *second-class* cause: The first-class Newtonian causes are forces (or the sources of forces), and what they cause, the first-class form of a Newtonian effect, is a change or deviation from an inertial state of motion. There is no doubt that for Newton, once the first law is in place, one can ask what causes the Earth to orbit the Sun (rather than travel at constant speed in a straight line), and that the cause in this case is the gravitational force produced on the Earth by the Sun. It is this sort of conceptual structure that allows us to so readily identify the cause of the motion in particle *P* above: Since the inertial state of *P* is for it to remain at rest, its change into a state of motion requires a cause, that is, a force, and the only force is provided by *Q*. Without the law of inertia, it would not be clear that the sudden onset of motion in *P* required any cause at all: Perhaps particles sometimes just spontaneously start moving. Or perhaps the inertial state of motion of a particle could be a jerky motion: The particle moves at a given velocity for a while, then stops for a while, then resumes at the same velocity, so the onset of motion in *P* is just the continuation of its natural state of motion.

Let us denominate laws *quasi-Newtonian* if they have this form: There are, on the one hand, *inertial* laws that describe how some entities behave when nothing acts on them, and then there are laws of *deviation* that specify in what conditions, and in what ways, the behavior will deviate from the inertial behavior. When one conceives of a situation as governed by quasi-Newtonian laws, then typically the primary notion of an effect will be the deviation of the behavior of an object from its inertial behavior, and the primary notion of a cause will be whatever sort of thing is mentioned in the laws of deviation.

Laws, of course, need not be quasi-Newtonian. At a fundamental level, the laws of the Game of Life are not. In that game, if a cell is occupied at one moment then it has no "natural" or "innate" tendency to remain occupied—or to become unoccupied—at the next moment. The patterns evolve in an orderly way, but the behavior of the overall pattern cannot be analyzed into, on the one hand, inertial states of the parts and, on the other, interactions that cause deviations from those states. There is no division of the rules of evolution into the inertial rules and the rules for change of inertial state.

That does not, of course, mean that the Life world is without causes. There are causes and effects, but the way we conceive of them, at the fundamental level, is quite different from the way we think of causes in a quasi-Newtonian setting. If a particular cell in the Life world (or, more obviously, the modified Life world) is occupied at a moment, then the cause of that is the complete 3-by-3 pattern of occupation centered on that cell the instant before, and similarly if a cell is empty. And the cause of that 3-by-3 pattern a moment before is the complete 5-by-5 pattern of occupation centered on that cell two moments before, and so on. As we go further back in time, the reach of the ancestral causes grows outward like a ziggurat, the Life-world analogue of the past light-cone of an event in a relativistic universe. At a fundamental level, the only proper causes in the Life world are these complete patterns at a time, which—in conjunction with the laws—generate the successive patterns that culminate in the given cell being occupied or empty at the given moment. (In the context of the discussion I am assuming that we agree that only a full 3-by-3 pattern is a cause of the succeeding state of the central square.)

This notion of causation will be available in any world governed by deterministic laws, whether quasi-Newtonian or not, and so the philosopher hankering after the most widely applicable concept of causation will likely be drawn to it. It is obviously an instance of the INUS concept of causation, for example. But I think that we fall back on this notion of causation only when circumstances demand it: Our natural desire is to think of the world in quasi-Newtonian terms, in terms of inertial behavior (or "natural" behavior) and deviations from inertial behavior: in terms of, to use a concept from mathematical physics, perturbation theory. Largely, this is because it is much easier to think in these terms, to make approximate predictions on the basis of scanty data, and so on. And often circumstances allow us to think in quasi-Newtonian terms even when the underlying laws are not quasi-Newtonian, or to think in macrolevel quasi-Newtonian terms quite different from the laws that obtain at the fundamental level. Indeed, the search for quasi-Newtonian laws does much to explain the aims of the special sciences.

4 Causation and Macrotaxonomy in the Special Sciences

I am a realist about laws: I think that there are laws, and that their existence is not a function of any human practices. I am also a primitivist about laws: I do not think that what laws there are is determined by any other, distinctly specifiable set of facts, and that in particular it is not determined by the total physical state of the universe. And I am a physicalist about laws: The only objective primitive laws I believe in are the laws of physics. Speaking picturesquely, all God did was to fix the laws and the initial state of the universe, and the rest of the state of the universe has evolved (either deterministically or stochastically) from that. Once the total physical state of the universe and the laws of physics are fixed, every other fact, such as may be, supervenes. In particular, having set the laws of physics and the physical state, God did not, and could not have, added any further laws of chemistry or biology or psychology or economics.

We do not, however, understand the vast majority of what we do understand by reflecting on the laws of physics. For example, there is much that I understand about how the computer I am now using works, and precious little of that derives from detailed knowledge of the physics of the machine. Rather, I understand its operation by thinking about it in terms of some *lawlike generalizations*, generalizations that resemble laws at least insofar as being regarded as supporting counterfactual claims and being confirmed by positive instances.

In this sense, it is a lawlike generalization about the computer that when it is turned on and the word processing program is running and there is document open, pressing a key on the keyboard will be followed by the corresponding letter appearing at the point where the cursor is, and the cursor will move over one space to the right (unless it is the end of a line, in which case it moves all the way to the left on the next line, unless it is at the bottom of the window, etc.). This generalization, which could be made more precise and extensive at the cost of much tedium, is taken to support counterfactuals: If I had hit the letter "*z*" on the keyboard instead of "*s*" just before the last colon, the word that would have appeared would have been "counterfactualz."

No doubt it is by means of such generalizations that we understand how to use the computer, predict how it will function, explain it to others, and so on. And no doubt this generalization, albeit lawlike in certain respects, is not in any metaphysically interesting sense a law. If I should hit a key and the corresponding letter did not appear, then it is not that any *law* would be broken: Rather, *the computer* would be broken (or misprogrammed, or crashed, etc.). And no doubt if the generalization is correct and the counterfactuals it implies are true, that is ultimately because of the *physical* structure of the machine operating in accord with the laws of physics. The lawlikeness of the macrogeneralizations, insofar as they are lawlike, is parasitic on the laws of physics in a way that the laws of physics are not parasitic on anything.

The point is that this is how the special sciences work: They seek to impose a taxonomy on the physical structure of the world (the concept "keyboard," for example, is not and cannot be reduced to the vocabulary of physics) in such a way that the objects as categorized by the taxonomy fairly reliably obey some lawlike generalizations that can be stated in terms of the taxonomy. Generalizations about how computers, or cumulus clouds, or volcanoes, or free markets behave are evidently of this sort.

Talk about "carving nature at the joints" is just shorthand for "finding a macrotaxonomy such that there are reasonably reliable and informative and extensive lawlike generalizations that can be stated in terms of the taxonomy," and the more reliable and informative and extensive, the closer we have come to the "joints." Again, I claim nothing particularly novel, or astounding, about this observation.

But if the foregoing analysis is correct, we are now in a position to add something new. We have already seen that certain forms of laws, namely, quasi-Newtonian laws, allow the identification of causes to be particularly simple and straightforward. So insofar as the special sciences seek to use causal locutions, it will be a *further* desideratum that the lawlike generalizations posited by the sciences be quasi-Newtonian. The special sciences, and plain common sense as well, will seek to carve up the physical world into parts that can, fairly reliably, be described as having inertial states (or inertial motions) that can be expected to obtain unless some specifiable sort of interference or interaction occurs. Or at least, those special sciences that manage to employ taxonomies with quasi-Newtonian lawlike generalizations can be expected to support particularly robust judgments about causes.

A few examples. We obviously understand computers in quasi-Newtonian terms: Depending on the program being run, there is an inertial state or inertial motion, and that motion can be expected to continue unless some input comes from the keyboard or mouse. The input is then an interfering factor, whose influence on the inertial state is specified by some "laws" (the program). We similarly understand much of human biology in quasi-Newtonian terms. The inertial state of a living body is, in our usual conception of things, to remain living: That is why coroners are supposed to find a "cause of death" to put on a death certificate. We all know, of course, that this macrogeneralization is only a consequence of the clever construction of the body and a lot of constant behind-the-scenes activity: By all rights, we should rather demand multiple "causes of life" for every day we survive. Nonetheless, the human body, in normal conditions, is sufficiently resilient that the expectation of survival from day to day is a reliable one, and the existence of the right sort of unusual circumstance immediately preceding death is typical. We do sometimes say that people just die of old age (which is obviously not a normal sort of cause) when no salient "cause of death" exists, and our acceptance of this locution illustrates our awareness that the quasi-Newtonian generalization "In the normal course of things (in the absence of 'forces') living human bodies remain alive" is not really a law at all.

Most critically for our present purposes, we typically think of the operation of *neurons* in quasi-Newtonian form: The inertial state of a neuron is not to fire, and it departs from that state only as a result of impinging "forces," namely, electrochemical signals coming from other neurons. These "forces" come in different strengths and can be either excitatory or inhibitory, and there is a lawlike formula that describes how the neuron will depart from its inertial state depending on the input to it. Of course, I do not mean to defend this conception as a bit of real neurophysiology, but that is how we lay people (and particularly we philosophers) tend to think of neurons. So conceived, it is often very easy to identify the cause of the firing of a neuron. If a neuron fires and only one excitatory impulse came into it, then that impulse was the cause of firing. End of story. Counterfactuals about what would have happened had that excitatory impulse not come in (whether, for example, some *other* impulse would have come in) are simply irrelevant. Details of the larger neural net in which these two are embedded are simply irrelevant. The only thing that would *not* be irrelevant would be the discovery that the quasi-Newtonian generalization is false, for example, because neurons sometimes spontaneously fire with no input.

The widespread use of Lewisian "neuron diagrams" in discussions of the nature of causation is, from this point of view, both a beautiful confirmation and a deep puzzle. We like neuron diagrams because our intuitions about what is causing what are strong and robust. Those who want to analyze causation in terms of counterfactuals think that the diagrams are useful as tests of counterfactual analyses: The trick is to find some condition *stated in terms of counterfactuals about firing patterns* that picks out all and only the causes. Those conditions then tend to get very complicated and quickly embroil one in issues like backtracking, miracles, intermediate states, and so on. But it is perfectly apparent that our strong and robust intuitions in this case are not generated by fancy considerations of counterfactuals at all: They are generated by the application of quasi-Newtonian *laws* to the situation, and the counterfactuals be damned. So the puzzle is why it has not been apparent that the very diagrams used to discuss counterfactual analyses have not been recognized as clear illustrations of the wrongheadedness of the counterfactual approach.

The great advantage of the special sciences *not* being fundamental is the latitude this provides for constructing macrotaxonomies well described by quasi-Newtonian generalizations. For example, one can try to secure the reliability of a quasi-Newtonian law of inertia by *demanding* that one find an interfering factor if the inertial state changes. In a case like finding the "cause of death" this can often be done. There is a lot of biological activity all the time, and even absences and lacks can count: One can die of starvation or suffocation. There is also latitude in carving the joints: One can shift boundaries so as to identify systems that obey more robust generalizations. This sort of boundary-shifting helps explain our causal intuitions in some cases that have been discussed in the literature.

Consider the following example by Michael McDermott:

> Suppose I reach out and catch a passing cricket ball. The next thing along in the ball's direction of motion was a solid brick wall. Beyond that was a window. Did my action prevent the ball hitting the window? (Did it cause the ball *not* to hit the window?) Nearly everyone's initial intuition is, "No, because it wouldn't have hit the window irrespective of whether you had acted or not." To this I say, "If the wall had not been there, and I had not acted, the ball would have hit the window. So between us—me and the wall—we prevented the ball from hitting the window. *Which one* of us prevented the ball hitting the window—me or the wall (or both together)?" And nearly everyone then retracts his initial intuition and says, "Well, it must have been your action that did it—the wall clearly contributed nothing." (1995a, p. 525)

McDermott's argument is quite convincing, but a puzzle remains. Why was nearly everyone's *initial* reaction the "wrong" one? What were they thinking? Was it merely the falsehood of the Hume counterfactual, "If you had not caught the ball it would have hit the window," that makes people judge that catching the ball is not a cause of the ball failing to hit the window? Then why does one *not* make this error (or not so easily) when the first catcher is followed by a second, infallible catcher rather than a brick wall?[4] The Hume counterfactual also fails here (if you had not caught the ball, the second catcher would have, and it still would not have hit the window), but it seems clear in this case that you are the actual cause here, and the second catcher merely an unutilized backup. The pair of examples makes trouble for any counterfactual analysis, since the counterfactuals in each case are identical, substituting the second catcher for the wall. So if we judge causes by means of counterfactuals, our intuitions should swing the same way in both cases, but they don't.

Here is an account that makes sense of the data. In judging causes, we try to carve up the situation into systems that can be assigned inertial behavior (behavior that can be expected if nothing interferes) along with at least a partial specification of the sorts of things that can disturb the inertial behavior, analogous to the Newtonian forces that disturb inertial motion. Let us call the things that can disturb inertial behavior "threats": They are objects or events that have the power—if they interact in the right way—to deflect the system from its inertial trajectory. We then think about how the situation will evolve by expecting inertial behavior unless there is an interaction with a threat, in which case we see how the threat will change the behavior. A threat can itself be threatened: Its inertial trajectory might have it interacting with a target system, but that trajectory be deflected by something that interferes with *it*. This is what we mean by *neutralizing* a threat, or by *preventing* an event.

Now what counts as inertial behavior, and what counts as a threat, depends on how we carve the situation up into systems. If we consider a normal window on its own, its inertial behavior is to remain unbroken (something must cause it to shatter, but nothing causes it not to if nothing interferes with it). The sorts of thing that count as threats are objects with sufficient mass and hardness and (relative) speed: A hurled rock is a threat to a window but a lofted marshmallow is not. The cricket ball in the example is obviously a threat to break the window (although the case just deals with hitting the window, the same point holds). The cricket ball is furthermore in a state of motion such that its (Newtonian) inertial trajectory has it hit the window, so for the threat to be neutralized, something must act to deflect the ball. This is clearly the first catcher in both cases, and the presence of the wall or second catcher never comes into the analysis.

But let us now carve up the situation a bit differently. Let us call the window plus brick wall system a "protected window." The inertial state of the window in a protected window system is to remain unbroken, and indeed to remain *untouched*. To count as a threat to this state, an object must be able to *penetrate* (or otherwise *circumvent*) the wall, at least if it is approaching from the direction protected by the wall. Thought of this way, the cricket ball is *not* a threat to the inertial behavior of the window; only something like an artillery shell would be. So in the given situation, there are no threats to the window at all, and a fortiori there are no threats that are neutralized, and so no prevention: *Nothing* causes the protected window not to be hit since that is its inertial state and nothing with the power to disturb the inertial state threatened.

In sum, if we carve things up one way (window plus wall plus catcher plus ball) we get systems governed by quasi-Newtonian generalizations that yield the judgment that the catcher prevented the window being hit; if we carve them up another way (protected window plus catcher plus ball) we get systems governed by quasi-Newtonian generalizations that yield the judgment that the window was never threatened with being hit, so nothing had to prevent it. And the interesting thing is that *both systematizations, together with the corresponding quasi-Newtonian generalizations, yield the same counterfactuals*. So no counterfactual analysis of causation has the resources to account for the disparity of judgments: Carving up the world differently can give different (special science) *laws*, governing different *systems*, but it should not give different truth values to counterfactuals.

If this analysis is correct, McDermott has played something of a trick on his test subjects. Their original intuition that the catcher does not prevent the window being hit is perfectly correct, relative to conceptualizing the situation as containing a protected window. McDermott then asks a series of questions that essentially require that one reconceive the situation, carving it instead into a window, a ball, and a compound wall plus catcher system. So thought of, the ball is clearly a threat to the window, the threat needs to be neutralized (or deflected) to prevent the window being hit, and the wall plus catcher system does the deflecting. The last step is to give the credit to either the wall or the catcher or both, and here the catcher clearly wins. All of this is perfectly legitimate: It is just no *more* legitimate that lumping the window and wall together and judging that nothing prevented the hit.

The window plus two catchers example tends to confirm the analysis: Here the natural tendency would be to regard the two catchers as equally autonomous systems. It would be odd to carve up the situation so that the window plus second catcher is regarded as a single system (although one could tell stories that would promote this). So our intuitions go differently from the wall case, even though the corresponding counterfactuals are the same.

The only remaining question is why typical subjects would have the tendency to regard the situation as containing a protected window, rather than in terms of a regular window, a threatening ball, and a pair of potential neutralizers. Lots of hypotheses come to mind, and the only way to test them would be to try many more sorts of examples. Windows and walls are obviously more similar than catchers and walls, since windows and walls are inert bits of architecture. It is *simpler* to keep track of things using the "protected window" systematization, since there are fewer threats and hence fewer threats that need to be tracked. The "protected window" systematization also has the advantage of yielding (in this case, at least) causal judgments that agree with the Hume counterfactual: The catcher does not prevent the window from being hit, and the window would not have been hit even if the catcher had not acted. (The last two considerations are held in common with the "two catcher" case, where our intuitions go the other way, but in that case there is positive pressure to group the catchers together.)

A thought experiment recommends itself: Imagine various ways for the wall to metamorphose into a second catcher and track when one's causal intuitions flip. But that is a task for another paper.

Alternative taxonomies need not be at the same degree of resolution, as in the foregoing example. Ned Hall has argued ("Two Concepts of Causation," chap. 9, this vol.) that cases of double prevention (preventing an event that would have prevented another event) are typically not regarded as causation (in the sense of production). Jonathan Schaffer (2000c) has then pointed out that guns fire by double prevention, but one is not tempted to conclude that pulling the trigger of a gun is not a cause of the firing. But surely what guides out intuitions in this case are lawlike macrogeneralizations in quasi-Newtonian form. The "inertial law" for guns is just: A gun doesn't fire if nothing acts on it (e.g., pulls the trigger, jars it, etc.). *Anything* that regularly results in a gun firing (particularly pulling the trigger) counts as a cause that changes the inertial state, no matter how the trick is done at the microlevel. *This judgment is not reversed even if one decides that the right thing to say at the microlevel is that pulling the trigger does not produce the firing*. If different taxonomies can allow for different lawlike generalizations and hence different causal judgments, we have all the makings for interminable philosophical disputes, since causal judgments can be reversed simply by changing taxonomies, as McDermott's example illustrates.

5 Remote Causation

The foregoing account has been concerned with the analysis of what we may call *proximate* or *immediate* causation: Where a situation is conceptualized as governed by quasi-Newtonian laws, the laws will specify what counts as an inertial state and therefore what counts as a deviation from an inertial state (a first-class effect), and also what sorts of things or events (causes) bring about such deviations. The interfering factor is then the proximate cause of the deviation, as in the billiards example.

This analysis does not, however, solve the problem of *remote* causation. We commonly identify events or actions as causes that are not the proximate causes of their effects: The loss of the horseshoe nail ultimately causes the loss of the battle not directly but by a long chain of intermediates. The loss of the horseshoe nail is only the proximate cause of the loss of the horseshoe.

Prima facie, the use of quasi-Newtonian laws appears to be of some help in this regard. If the laws have the form of specifying inertial states and interfering factors, then any concrete situation can be represented by what we may call an *interaction diagram*, a graph depicting inertial motions of objects as straight lines and divergences from inertial motion as always due to the interaction with some interfering element. One might think, for example, of Feynman diagrams as examples of the form. Note that if all interactions are local, or by contact, then interaction diagrams will look like spacetime diagrams, and the lines will represent continuous processes in spacetime, but this is not essential to the form: if there is action-at-a-distance, then two lines can intersect on an interaction diagram even if the objects they represent never come near each other in spacetime.

Interaction diagrams supply one simple method for identifying remote causes: Trace the diagram backward from an event, and every node one comes to counts as a cause. But this method will not, in many cases, accord with our intuitions. Being in the interaction diagram for an event may well be a necessary condition for being a cause, but it is unlikely to be sufficient. Interaction diagrams will include events that we consider not to be causes (in the usual sense) but failed attempts at prevention, such as the futile course of chemotherapy that fails to stem the cancer. Interaction diagrams also stretch back indefinitely in time, to events too remote to be commonly considered causes. They are a bit more parsimonious than the entire back light-cone of an event, but not that much.

This sort of situation is, of course, the siren song for the analytic philosopher. Perhaps we can identify commonsense remote causes by use of an interaction diagram plus some further condition (the Hume counterfactual, for example). But the avalanche of counterexamples to the many theories of causation that have been floated ought to give us pause. Perhaps there is some reason that the analysis of remote causation has proved so difficult.

So let's step back a moment and ask what *purpose* there is to identifying remote causes in the sort of situation we are contemplating, that is, where we know the laws, the immediate causes of events, and the complete interaction diagram. What more could we want? What do we gain by trying to distinguish events that count as remote causes from others that are not?

One thing we might have is a practical concern for prediction and/or control. We might like to know how we could have prevented a certain event, or whether a similar event is likely to occur in future situations that are, in some specified respects, similar. But in that case, all we would really care about is the Hume counterfactual —and that, as we know by now, is neither sufficient nor necessary for causation. So what other aim might we have?

In some circumstances, we are interested in remote causes because we wish to assign *responsibility* for an event, for example, when offering praise, or assigning blame, or distributing rewards, or meting out punishments. In these circumstances, identifying a remote cause is often tantamount to establishing a responsible agent, and many of our intuitions about remote causes have been tutored by our standards of culpability. If this is correct, then we might make some progress by reflecting on such standards.

The game of basketball provides an example where rules for assigning credit as a remote cause have been made reasonably explicit. Consider this case: A pass is thrown down court. The center leaps and catches the ball, then passes it to Forward *A*, who dunks the ball. Forward *A* gets credit for the field goal, and the center gets an assist: He counts as a remote cause of the points being scored. Note that these attributions do not change even if it is also true that the pass was intended for Forward *B*, who was standing behind the center, and that had the center not touched the ball, Forward *B* would have caught the pass and scored, or would have passed it to *A*. Even if the Hume counterfactual does not hold for the center (even had he not caught the ball, the points would have been scored), the center counts, unproblematically, as a cause.

Now consider an apparently analogous case. John enters a car dealership, unsure whether he will buy a car. He is met by an Official Greeter, who directs him to Salesman *A*. Salesman *A* makes the pitch and convinces John to buy the car. Had the Greeter not intercepted him, John would have run into Salesman *B*, who also would have convinced him to buy the car (or who, perhaps, would have directed him to Salesman *A*). In this case, the Greeter can not, intuitively, claim credit for the sale: He did not even remotely cause John to buy the car. He was a cause of the cause, a cause of John's hearing Salesman *A*'s pitch (at least in the scenario where the alternative is a pitch from Salesman *B*), but we are not inclined to accept transitivity here.

Doubtless there are perfectly good reasons for the difference in practices for assigning credit in these cases: *Typically*, in a basketball game, had the assist not occurred points would not have been scored, so one wants to recognize and encourage those who give assists. But *typically*, in the sort of situation described in the car dealership, whether a sale is made does not depend on the actions of the Greeter. But what is typical does not affect the individual case: We can make the counterfactual structure of this *particular* pair of examples as analogous as we like (the schematic "neuron diagrams" for the scenarios can be made identical) without changing our views about who deserves credit and who does not. If so, then *no* generic account of remote causation couched in terms of interaction diagrams or counterfactual structure will always yield intuitively acceptable results: In different contexts, our intuitions will pick out different nodes on the same diagram as remote causes. In these cases, the definition of a remote cause is, as Hume would put it, "drawn from circumstances foreign to the cause," that is, from statistical features of other cases that are regarded as similar. As such, the standards violate the desideratum that causation be intrinsic to the particular relations between cause and effect.

If standards for identifying remote causes vary from context to context, and, in particular, if they depend on statistical generalities about *types* of situations rather than just on the particular details of a single situtation, then the project of providing an "analysis" of remote causation is a hopeless task. We might usefully try to articulate the standards in use in some particular context, but no generic account in terms of interaction diagrams or counterfactual connections will accord with all of our strongly felt intuitions.

6 The Metaphysics of Causation

If the foregoing analysis is correct, then (1) what causes what depends on the laws that govern a situation; (2) judgments of causation are particularly easy if the laws have quasi-Newtonian form; (3) everyday judgments ("intuitions") about causation are based not on beliefs about the only completely objective laws there are (viz., physical laws) but rather more or less precise and reliable and accurate lawlike generalizations; (4) the same situation can be brought under different sets of such generalizations by being conceptualized differently, and those sets may yield different causal judgments even though they agree on all the relevant counterfactuals.

To what extent, then, is causation itself "objective" or "real"? At the level of everyday intuition, the freedom to differently conceptualize a situation implies that one's causal judgments may not be dictated by the complete physical situation per se. Further, the lawlike generalizations appropriate to the conceptualization can be criticized on objective grounds: They could be more or less accurate or clear or reliable. If, for example, windows sometimes spontaneously shatter (as a result of, say, quantum fluctuations), then the reasoning that the ball made it shatter because the ball was the only thing to hit it (and its inertial state is to remain unbroken unless something hits it) is no longer completely trustworthy, the less so the more often spontaneous shattering occurs. Lawlike generalizations are also supposed to support counterfactuals, but those counterfactuals must ultimately be underwritten by physical law, so a close examination of the physics of any individual situation could undercut the macrogeneralization applying to that case. The quest for greater scope, precision, and reliability of generalizations tends to force one to more precise microanalysis, ultimately ending in the laws of physics, which brook no exceptions at all.

If the laws of physics turn out to be quasi-Newtonian, then there could be a fairly rich objective causal structure at the fundamental level. But if, as seems more likely, the laws of physics are not quasi-Newtonian, then there may be little more to say about physical causation than that the entire back light-cone of an event (or even the entire antecedent state of the world in some preferred frame) is the cause of the event, that being the minimum information from which, together with the laws of physics, the event can be predicted. Quasi-Newtonian structure allows one to differentiate the “merely inertial” part of the causal history of an event from the divergences from inertial states that are paradigmatic effects, but without that structure it may be impossible to make a principled distinction within the complete nomically sufficient antecedent state. In that case, causation at the purely physical level would be rather uninteresting: If all one can say is that each event is caused by the state on its complete back light-cone, then there is little point in repeating it. None of this, of course, is of much interest to physics per se, which can get along quite well with just the laws and without any causal locutions.

There is much more that needs to be said about the role laws, or lawlike generalizations, play in the truth conditions of causal claims. The proposal I have made is more of a sketch than a theory, and it has all the resources of vagueness to help in addressing counterexamples and hard cases. Perhaps there is no adequate way to make the general picture precise. But the counterfactual approach to causation has had a good long run, and it has not provided simple and convincing responses to the problem cases it has faced. Perhaps it is time to try another approach.

Notes

1. Ned Hall has suggested that there are really two concepts of causation, one of which (“dependence”) amounts to nothing more than the Hume counterfactual, whereas the other (“production”) cannot be so

analyzed (cf. his "Two Concepts of Causation," chapter 9 in this volume). If one wishes, this essay can be read as an account of production, although I suspect that at least many cases of dependence can be covered using these methods.

2. Natural law can be employed in analyses of causation in many ways. One interesting, and distinct, approach is advocated by Schaffer (2001) in his PROPs analysis: Schaffer uses laws to identify the process that actually brought an event about.

3. If one imposes some natural symmetry constraints, such as 90° rotational symmetry, then there will be fewer than 512 distinct cases to deal with.

4. This example is discussed by John Collins in "Preemptive Prevention," chapter 4 in this volume.