

CAUSATION

At one time or another, most of us have had the experience of reaching for something too quickly. Let's say you go to grab some bread from across the dinner table and bump a glass with your elbow, spilling water all over the table. Nothing could be more obvious, it seems, than that you bumped the glass and *caused* the water to spill. To put this in a grand-sounding way, you stood in the relation of *causation* to the spilling water. That is the concept that we are going to try to better understand in this chapter, the relation that seems to hold between you and the water in this mundane example. This will be our first in-depth look at what is primarily a conceptual issue in metaphysics. It is a terrific place to begin because causation is about as familiar, central, and tricky as concepts come.

The spilled-water case should be enough to convince you that causation is a very familiar concept. That it is a very central concept is almost as obvious: It is what we do whenever we affect the things around us. It is what those things do to us whenever they affect us. Molecular bonding, planetary rotation, human decisions, and life itself are all causal processes, they all somehow involve causation. Causation is part of scientific practice: At least typically, a scientific explanation of some event will include some mention of something that caused that event; you can't say why something happened without identifying something that caused it to happen. Causation is part of philosophy, too. As this very chapter makes clear, it is part of philosophy because philosophers try to understand it better. More interestingly, it also plays a role in philosophy when the focus is on other matters. As we saw in Chapter 2, philosophers doing ontology wonder how we could know anything about abstract entities like numbers if it is true that none of these entities cause anything. As we will see in

Chapter 5, the most interesting arguments about the relationship between the mind and the brain turn on how non-physical mental states could affect the physical world.

Why do we say that causation is tricky? As we are about to see, despite its familiarity and its centrality, there is a good deal of disagreement among philosophers about how to understand causation better. But, causation's trickiness amounts to more than just that. (After all, disagreement is common in philosophy.) The further point we would like to make is that there is something about causation that makes providing a plausible theory of causation difficult. So much so, that unlike for many other areas of metaphysics, the current philosophical literature on causation is to a large degree not focused on which of two basic theories is correct. There is nothing corresponding to dualism vs. materialism (about the mind), compatibilism vs. incompatibilism (about freedom of the will), presentism vs. eternalism (about time), or endurantism vs. perdurantism (about identity). No, the study of causation has come to be better defined by a range of clever examples than by different stances one could take about causation. The main goal of this chapter will be to present these examples.

1. Preliminaries

There are some preliminary matters that need to be addressed before getting to the examples. We need to limit our attention to a certain range of sentences that use the word 'caused', a range of sentences that clearly express the relation about which philosophers have been most concerned. Something will also need to be said about what kinds of entities this relation is said to relate.

No doubt you have seen or at least heard the U.S. Surgeon General's warning: *Smoking causes lung cancer*. That statement is not describing any particular smoking event. It does not explicitly mention any particular person, place or thing at any particular time doing any causing of anything. In this way, this claim and others like it are different from the claim that you caused the water to spill, which, in our opening example, was very much about what you did at a particular time and place. It may be better to think of the report about smoking as describing a causal relation between universals. Maybe the Surgeon General is saying that the property of smoking causes the property of having cancer. Or, it may be that the report needs to be understood as some kind of generalization about individual cases. Maybe it is saying that everyone who smokes gets cancer. That is probably too strong, but maybe the Surgeon General is making a weaker generalization, just saying that the number of cases where smoking causes cancer is high. In any case, this sort of causation—if that's what it is—is sometimes called *property-level* or *general-case* causation. It is not the focus of this chapter.

Instead, we will look at straightforward cases where both the cause and the effect are pretty clearly particulars of some kind. Philosophers sometimes think of this as a kind of causation and label it *single-case* causation. Reports of single-case causation are expressed by the following sentences:

- (1) You caused the water to spill.
- (2) The eruption of Mt. Vesuvius caused the destruction of Pompeii.
- (3) The iceberg caused the Titanic to sink.

- (4) Joe's pulling the trigger caused the gun's firing.

- (5) The Red Sox lost the 1986 World Series because the ball went through Bill Buckner's legs.

Since our sole concern will be with single-case causation, we'll just call it causation.

Even restricting our attention to causation so understood, there is still some more sorting out to be done. Some philosophers will deny that Sentence (5) is a causal sentence. They'll say that 'because' is not a proper word to express causation. They'll claim that 'because' sentences really express *explanation*, and maintain that explanation is different from causation. We are suspicious of that sort of position because Sentence (5) seems to us just to be saying that the ball's going through Bill Buckner's legs caused the Red Sox to lose the Worlds Series. But, to keep things simple, we will set sentences like (5) aside. There are, however, still some differences among the remaining four kinds of sentences that shouldn't be overlooked. Sentence (1) says that *you* (a person) caused the water to spill. Sentence (2) says that *an eruption* (an event) did some causing. Sentence (3) has it that *the iceberg* (a material object) sunk the Titanic. Sentence (4) says that *Joe's pulling the trigger* was the cause. Some philosophers will take that to be an event also, others take it be a *state of affairs* or a *fact*. (We won't go through it, but we get similar variation with respects to the ontological character of the effects mentioned in (1)-(4), too.) Theses differences have led to a remarkable number of philosophical reactions. Some philosophers have been inclined to think that there is more than one (single-case) causal relation. They think, for example, that material object causation is not event causation and that these two relations call for different metaphysical accounts. Others have thought

that there is just one causal relation, but that it can relate many different kinds of things. Some try to paraphrase certain causal sentences in such a way that all causal sentences can be seen at a fundamental level to involve one relation that relates only one kind of thing. Many philosophers think that causation is most fundamentally a relation between events. But even they don't agree on what an event is! Some of them think events are unstructured, concrete particulars; others don't.

We will do our best not to get involved with these disputes. In presenting some analyses of causation, we will take, '*c* caused *e*' as our locution for analysis,¹ and try to minimize our assumptions about '*c*' and '*e*'. When we have to categorize '*c*' and '*e*', we will tend to follow what we take to be the norm and call them events. We do all this just to regiment our own discussion. It is not meant to reflect any important assumptions of our investigation. With that in mind, we will also work very hard not to make any special assumptions about what events are. Pretty much, if not all, of what we will say about events could equally well be said in only slightly different terms about states of affairs or

¹We will not be able to resist using locutions like '*caused x to F*' whose logical form is not at all obvious but is a common and natural way of formulating causal claims. We will also not make a big deal about the difference between sentences that use '*caused*' and those that instead use the phrase '*was a cause of*'. Uttering Sentence (2) somehow makes the suggestion that the eruption of Mt. Vesuvius was the *only* cause of the destruction of Pompeii, whereas saying, '*The eruption was a cause of the destruction*' almost seems to make the suggestion that *there were other* causes. This difference is not important to the study of causation. How could something have been *a cause of* something without also having *caused* that something? How could something have *caused* something without also having *been a cause of* that something?

facts, whether events, states of affairs, or facts are concrete or abstract things, structured or unstructured things, whatever. Just keep in mind that ‘*c*’ and ‘*e*’ always stand for a particular of some sort.

2. Theories of Causation

It goes without saying that a cause causes an effect only if both the cause and the effect occur. There is no causation linking a nonoccurrence with anything. So, any theory of causation subscribes to *c* caused *e* only if *c* and *e* both take place. It really does go without saying because, if an event didn’t occur, then the event wouldn’t even exist. That causes and effects must occur in order to be causes and effects is about the only thing that is not controversial about causation.

One idea that has played a central role in the history of philosophy since David Hume is that causation is constant conjunction. Hume said, “We may define a cause to be an object, followed by another, and where all objects similar to the first are followed by objects similar to the second” (1955, p. 79). Spelling out similarity in terms of sharing a property and putting this in terms of events *c* and *e*, we have our first theory for consideration:

Causation in terms of Constant Conjunction

c caused *e* if and only if there are properties *F* and *G*, such that *c* has *F* and *e* has *G*, and each thing of kind *F* is always followed by something of kind *G*.

As we will see in a moment, this is a pretty simplistic theory. It does, however, get some cases right. As illustration, consider Pompeii and the volcano. A little before Pompeii was destroyed, an eruption began. The eruption had a certain complex property. The eruption was such and such distance from Pompeii, involved a certain massive amount of lava that had a certain high average temperature and that was heading with a specific flow rate toward Pompeii. We would need to fill this in more, mentioning certain features of Pompeii, like its expanse, but, once filled in enough, it would be plausible to think that whenever anything had that complex property, the destruction of the nearby city would soon follow. So, arguably, according to the constant-conjunction analysis, and plausibly enough, it is true that the eruption of Mt. Vesuvius caused the destruction of Pompeii.

It is doubtful that Hume or any other philosopher had anything quite as simplistic as the constant-conjunction analysis just stated in mind as a full-fledged, final philosophical account of causation.² For instance, it might be true just as a matter of pure coincidence that every coin that has ever been in a particular brand new pair of pants has been a nickel. In fact, it may be that there was just one such coin and that is the only coin that will ever be in the pockets of those pants; the pants will be destroyed in a fire tonight. So, according to the constant-conjunction analysis, the coin's being in those pants caused it to be a nickel; anytime any coin has the property of being in those pants, its being in those pants is followed by its being a nickel. Of course, that is not at all a plausible or intuitive consequence of the constant-conjunction analysis. Any theory of causation that gives us a result like this needs to be reworked or rejected. To avoid counterexamples of this sort, proponents of a constant-conjunction analysis will insist that the constant conjunction has to be a constant

²Beauchamp and Rosenberg (1981) devote their book to spelling out the sophistications needed to make a recognizably Humean account of causation plausible.

conjunction of the right sort. Not just any true regularity connecting one kind of event to another can underwrite causal truths. Traditionally, proponents of a constant-conjunction analysis have argued that only laws of nature (or certain sorts of laws of nature) have that status.

Just so, a popular descendent of the constant-conjunction approach is an account defended by J. L. Mackie (1974). (Also see Bennett 1988.) Greatly simplified, the idea is that what is important to causation is that the cause be an *NS condition*, that the cause's occurring be a *necessary* part of a condition that together with the laws of nature is sufficient for the effect to occur.

Causation in terms of NS conditions

c caused e if and only if c is an NS condition for e .

So, consider Mt. Vesuvius's erupting and Pompeii's being destroyed. Setting aside for the moment that the laws of nature of our universe may be indeterministic in important respects, it is plausible to think that there were conditions of our world at the time of the eruption that, in conjunction with the fact that Mt. Vesuvius erupted, together with the laws of nature, entail that Pompeii would be destroyed. What's more, without the fact that Mt. Vesuvius erupted, those conditions together with the statement of laws of nature would not entail that Pompeii would be destroyed. The idea is that the eruption was a really important part of a certain time-slice of the universe that under the governance of the laws of nature led ultimately to the destruction. With the eruption taking place, there was enough going on so that the destruction had to take place; without the eruption taking place, there wasn't.

In 1973, David Lewis made popular and sophisticated a different idea also to be found (briefly) in Hume. It is somewhat similar to the Mackie-style approach in that both ideas are based on the thought that the cause has to somehow be necessary for the effect. Suppose that I am standing alone near an old abandoned house and that I whip a baseball at one of the windows. The window shatters. Evidently, my throw caused the window to shatter. But it also seems perfectly true that, if I hadn't thrown the baseball, then the window wouldn't have shattered. Maybe, causation between two events just amounts to its being the case that, if one event had not occurred, then the other would not have occurred.

Causation in terms of Counterfactual Dependence

c caused *e* if and only if, if *c* hadn't occurred, then *e* wouldn't have occurred.

In a straightforward way, this analysis avoids the very basic problem offered above as trouble for the constant-conjunction analysis. A more interesting relation is required to hold between the events than just mere constant conjunction. It is not true that, if the coin weren't in those pants, then it wouldn't have been nickel. Clearly, it still would have been a nickel; it just would have been a differently located nickel. Our other stand-by cases are readily handled too. For example, if the eruption of Mt. Vesuvius hadn't occurred, then neither would have the destruction of Pompeii.

Be aware: Our analysis of causation in terms of counterfactual dependence is not an analysis Lewis ever adopted, though he did think that counterfactual dependence between distinct events was a sufficient condition for causation. We will consider other aspects of Lewis's more fully developed position as the chapter goes on. Specifically, we will consider a more sophisticated counterfactual

analysis advanced by Lewis that has similarities to other theories of causation that invoke probability due to concerns about the possibility that the world is indeterministic. We will also briefly consider revisions to our counterfactual analysis and the Mackie-style analysis involving causal chains of intermediate events. Of course, many other analyses of causation have been offered. Indeed, there are so many theories and variations of theories dating back so far that there is no hope of covering all of them or even an important survey of them in a single chapter. The most important two classes of such theories getting the shortest shrift are the transference theories and the manipulability theories.³ Transference theories say that causation always involves the transfer of something (e.g., momentum) from one object to another. The main challenge for these theories is to say what is transferred from one thing to another in *all cases* of causation. Manipulability theories say that one event caused another if and only if the first event would have been an effective means for bringing about the other. The main challenge here is make clear why this sort of account is not doomed to circularity, since bringing about and causing seem to be the same thing.

Rather than saying anything more right now about how philosopher have theorized about causation, we will cover an important series of core examples that any plausible and suitably interesting theory of causation will need to say something about. It is our belief that a student who knows these examples probably gains a better understanding of causation and is certainly better prepared to embark on evaluating new theories of causation than a student that only studies the theories.

³Recent work: Ehring (1997) defends of a trope-transference theory. Menzies and Price (1993) defend a manipulability theory.

3. Three Core Examples: Deterministic Causation

In this section, we will introduce three core examples that all bring out some interesting features of causation and also raise at least preliminary questions for our Mackie-style analysis and our counterfactual analysis. These three examples, and the three more examples to be introduced in Section 4, make perfect sense and are just as effective even if we assume that our world is a deterministic world. Three further core examples only make sense under the assumption of indeterminism. These have prompted major and important revisions to analyses of causation and raise lots of new issues, and will not be introduced until Section 5. All the cases will be raised relatively quickly, with only some hints as to how they might be addressed. What's important is familiarization with the cases.

a. Noncausal Connections

Actually, we considered one case of a noncausal connection in Section 2, that of the nickel in the new pair of pants. It presented a serious problem for the constant-conjunction analysis because being a coin in that pocket of those pants correlates with being a nickel, though there is not a corresponding causal connection between the coin's being in that pocket of those pants and that coin's being a nickel.

A different kind of noncausal connection arises with *epiphenomena*. Let's start this example by making a certain simplifying assumption: that it is true and a law that, whenever a properly functioning barometer's reading drops, a storm occurs shortly thereafter.⁴ Then suppose we have a very ordinary situation. A storm is approaching your hometown, the atmospheric pressure drops dramatically, and your properly functioning barometer records the drop in pressure. You predict there will be a storm. Your prediction turns out true. Your hometown is deluged. In this example, there is a constant conjunction linking changes in the barometric readings and storms. That obviously presents a problem for the constant-conjunction approach. It will have the crazy consequence that the change in the barometric reading I used to predict the storm caused the storm, and that's not correct. Barometers aren't weather-making machines! The theory needs to be revised or rejected. This is a case of epiphenomena because the change in the barometric reading is a secondary phenomena not really involved in the processes leading up to the storm striking your hometown. This concept of epiphenomena is an important one. In the study of disease, it is crucial for doctors to be able to distinguish the causes of the disease from symptoms. That is not always easily done.

This new problem for the constant-conjunction analysis is not avoided by making the move mentioned earlier; requiring that there be a law of nature connecting properties of the cause with some property of the effect will not help with this case. It is built into this new example that there is the corresponding law. (One might maintain that it is not the right kind of law, but that is not helpful unless *right kind* is defined in some way that does not make the analysis circular.) The Mackie-style

⁴Even with the properly-functioning clause included, that is probably not really a law—the barometer could be functioning and be inside a pressure cooker—but philosophers have always granted that for the sake of the example.

approach is faced with a similar problem. The change in the barometer reading together with the laws of nature entail that the storm will arrive. So, trivially, it is a necessary part of that sufficient condition for the effect. So, *prima facie*, the Mackie-style approach also has the unwanted consequence that the change in the reading caused the storm to arrive. For the counterfactual approach, everything depends on what would be the case if the reading hadn't dropped. If there still would have been the actual drop in atmospheric pressure (it is just, say, that the properly functioning barometer would be malfunctioning), then the storm still would have occurred. So, the counterfactual approach would give the right answer; the change in reading did not cause the storm. But is that what would happen if the reading hadn't shown a drop in pressure? Maybe, instead, there wouldn't have been a drop in pressure—the barometer would have kept working perfectly, it's just that there wouldn't have been a drop in pressure and so there also wouldn't have been a storm! On this way of evaluating the key counterfactual, it turns out true that the change in reading did cause the storm, and that's not correct.

A third kind of noncausal connection arises because of certain necessary connections between events. When Socrates drank the hemlock and died in the Athenian prison, his wife, Xanthippe, became a widow. (See Kim 1974, pp. 41-42.) Did Socrates' death cause Xanthippe to become a widow? It is clear that, had Socrates not died, Xanthippe would not have become a widow. It is also clear that Socrates' death is an NS condition of Xanthippe's becoming a widow. So, our counterfactual analysis and the Mackie-style analysis both count Socrates' death as a cause of Xanthippe's becoming a widow. Many philosophers find this conclusion unacceptable. The worry is that any relation holding between these two events is bound not to be of the tangible/physical sort one normally expects causation to be. Sometimes it is argued that, since Xanthippe was not in the prison with Socrates, his death could not have caused her to become a widow. She became a widow the

instant he died. Since there is no instantaneous causal action at a distance in a world like ours where no signals travel faster than light, his death didn't cause her widowhood.

b. Simultaneous Causation

There is an important part of the constant-conjunction analysis that we have not said much about. It requires that the events of kind F are always followed by—always occur before—the events of kind G . Hume also talked about the effect having to succeed the cause. Probably, you were just assuming that causes always occur before their effects, and that this was an implicit part of the Mackie-style and counterfactual analyses, too. Well, they are not. Bringing in time to help mark off which event is the effect and which event is the cause is actually a contentious part of the constant-conjunction approach.

It looks like a mistake to require that the time of the cause be prior to the time of the effect. Arguably, there are cases of causation between simultaneous events that would be ruled impossible by such a requirement. For example, suppose there is a perfectly rigid seesaw—when one end of the bar moves up or down, the other end moves in the opposite direction simultaneously. Also, suppose that you push down on one side and the other side goes up. Then the one side's going down caused the other side to go up. It is no objection to this example to point out that such a perfectly rigid seesaw is unrealistic. It may well be: That there is a perfectly rigid seesaw contradicts the law of nature that no signals travel faster than the speed of light. What really matters though is that the case is possible, and it sure seems possible. Furthermore, denying that there is simultaneous causation in

this simple case would be at odds with some of the good intuitions motivating the analyses set out in Section 2. For instance, consider the counterfactual analysis. If the side you pushed hadn't gone down at the time it did, then the other side would not have gone up at that very same time. So, without an additional condition ruling out the possibility of simultaneous causation, this analysis would imply that such causation is present, and intuitively that seems right. Thus, even by the lights of the counterfactual analysis, the causation should be there.

This is not to say that either the counterfactual analysis or the Mackie-style theory are not threatened by the case of simultaneous causation. They do appear to have the intuitive consequence that the side you pushed down on simultaneously caused the other side to rise, but they also seem to say that the other side's going up simultaneously caused your side to go down! That's the trouble for these two theories. The Mackie-style and the counterfactual analyses say that the seesaw case is a case of mutual causation. That doesn't seem right; *c* caused *e* but not *e* didn't cause *c*.

c. Causes vs. Background Conditions

Nelson strikes a match at his neighborhood cookout in order to light the coals beneath his grill. It is a beautiful sunny day, there are no rain drops to worry about, not a bit of wind. Not surprisingly, the match lights.

According to the standards set by the other examples from this chapter, this story of Nelson and his match should seem to be a pretty boring case. Indeed, it is. In fact, any ordinary case of causation would have sufficed to raise the present issue, which concerns whether there is any

important metaphysical difference between what are very clearly causes of the match's lighting and what are sometimes thought of as background conditions. Most everyone will agree that Nelson's striking the match caused it to light. Most everyone would agree that it would be very odd in typical conversations about the cookout to say, 'The presence of oxygen caused the match to light'. Did the presence of oxygen cause the match to light? Or was it only a background condition?.

This is an especially pressing issue for the Mackie-style analysis and the counterfactual analysis we are considering. Both of these views count lots of events as causes no matter what the effect in question is. Yes, they count the presence of oxygen as a cause of the match's lighting; the presence of oxygen is an NS condition for the match's lighting, and it is certainly true that, if oxygen hadn't been present, then the match wouldn't have lit. Counting the presence of oxygen as a cause is a conclusion that isn't so bad. There is a plausible explanation of why it would be odd to say, 'The presence of oxygen caused the match to light', a story based on the fact that participants in a typical conversation about the cookout will already take for granted that oxygen is required for a fire and that we shouldn't say what is already accepted as true. But these accounts also count other things as causes that it would be odd even to describe as background conditions. (We will focus on the counterfactual account but all the same points could be made about the Mackie-style approach.) If the Big Bang had not occurred, then the match wouldn't have lit. If Nelson hadn't been born, the match wouldn't have lit. If the matchbook were disintegrated by Martian X-rays, the match would not have lit. If you had swiped the matchbook from Nelson, it would not have lit. So, on the counterfactual account, not only is the oxygen a cause—so is the Big Bang, Nelson's birth, that no ray guns from Mars blasted the matchbook, and that *you* didn't swipe the matchbook.

4. Three More Core Cases: Deterministic Causation

Mostly to give the reader a break by making sure that Section 3 wasn't too long, we have saved three examples involving deterministic causation for this fourth section. These cases also have a lot in common with each other, including getting a lot of attention in the philosophical literature. Much of the attention can be attributed to the fact that Lewis found this group of problems, especially the preemption case, to be a serious challenge to his counterfactual analysis. Let's get too them.

a. Overdetermination

The story here is the story of a deserter sentenced to death. To keep it simple, we will suppose that his firing squad includes just two members, A and B, each of whom is a crack shot. The commanding officer gives the order and the sharpshooters fire their loaded rifles at exactly the same time. Side by side, the bullets from the two rifles simultaneously pierce the center the heart of the deserter, who dies soon thereafter from massive blood loss. Each shot was such that it would have caused the deserter to die even if the other shot had never been fired. In some ways, this is a simple case. In other ways, it is not. It is certainly simple enough to describe the case as we just did. The hard part is coming to a reasonable conclusion about what the causal truths are about this situation and what this says about our theories of causation.

It is tempting to think that each of the shots caused the deserter to die. Given the circumstances at the time the triggers were pulled, given the laws governing how events take place

in our example, each shot was an NS condition for the death. It is even part of the description of the example that either shot, had it been the only shot fired, would have caused the death. So, it would seem strange if somehow they could lack the effects they would have had individually in virtue of both occurring. These considerations, if telling, would stir up a whole lot of trouble for our counterfactual analysis. On a perfectly natural way of evaluating the relevant counterfactual, it seems that, if A hadn't fired, then the deserter still would have died. So, according to our counterfactual analysis, A's shot didn't cause the death. (The same point could be made about B.) But that goes directly against the tempting conclusion that each shot was a cause of the death. The same problem doesn't arise for the Mackie-style approach; it counts both A's shot and B's shot as causes of the death.

But hold on. Don't dismiss our counterfactual analysis too quickly! This is not a case where our intuitions are very clear or very strong. Maybe the counterfactual theory has things right. Notice that A's shot didn't really make any difference regarding the deserter's death. The sharpshooter would have died even if A had felt a last second twang of remorse and hadn't shot. Maybe all that's true about the case is that the deserter died because *at least one* of the shots was fired. If that's the right description of where the causation is in this case, then the counterfactual account seems to get things exactly right. If at least one shot hadn't been fired, the deserter would have lived. The Mackie-style approach will agree that at least one shot being fired caused, but will hold to the now-not-so-obvious conclusion that it is also true that A's shot and that B's shot each caused the death.

b. Preemption

Let's take a different kind of case where there is more agreement among metaphysicians about the causal facts. Two live wires, one from Switch A and one from Switch B, lead through a junction box to a light bulb. Both switches are currently open, so no current reaches beyond the switches. The junction box is interesting because it only lets the current from one wire through at a time. More specifically, it will allow current to pass through from the first switch that is closed but not the second. In the case of a tie, only the current from Switch A passes through. Now what happens in the case of interest here is that both Switch A and Switch B are closed at the same time, the current from Switch A passes through the junction, it gets to the light bulb, and the light bulb goes on. The current from Switch B stops once it gets to the junction box. What do you think the causal facts are about this hypothetical situation? Did the fact that Switch A was closed cause the light to go on? Did the fact that Switch B was closed cause the light bulb to go on? Pretty clearly, it was Switch A that was responsible.

This is what philosophers call a case of *preemption*. They call it that because there are two potential causes (in our case, the closing of Switch A and the closing of Switch B), but one of these events is preempted by the other from bringing about the effect. In some ways, preemption cases are like cases of overdetermination. For example, if either Switch A or Switch B hadn't been closed, the light bulb still would have gone on, just as if either Sharpshooter A or Sharpshooter B hadn't fired, then the deserter still would have died. So, as in the overdetermination case, neither of our two potential causes makes a difference with respect to the effect. (Maybe neither Switch A nor Switch B is a cause here?) In other ways, preemption cases are a little different from cases of

overdetermination. In typical overdetermination cases, there is a certain symmetry associated with the two potential causes that is not present in preemption cases. Sharpshooter A and Sharpshooter B have exactly the same claim to being causes of the deserter's death. Not so, for Switch A and Switch B. Regarding what causes what, Switch A seems to have a lot more going for it. The current makes it all the way from Switch A to the light bulb. No current makes it all the way from Switch B to the light bulb. For this reason, the widely accepted take on this preemption case is that Switch A caused the light bulb to go on and that Switch B did not have that same effect.

Preemption cases have been a thorn in the side of a wide-range of philosophical theories about causation. The counterfactual analysis has the consequence that closing Switch A is not a cause of the light bulb going on; if Switch A hadn't been closed, then the light still would have gone on. The problem for the Mackie-style account is different. It correctly counts that Switch A was closed as a cause of the light going on; the problem is that it counts that Switch B was closed also as a cause of the light going on. The closing of Switch B is an NS condition of the light going on.

The standard way for these theories to try to sidestep their problems with the preemption case is to make use of some intermediate chain of events running between the cause and the effect. The analysis Lewis offered in 1973 identified causation with the ancestral of counterfactual dependence. (See 1973, p. 563.) That is a fancy way of saying that an event could get counted as a cause *either* by having the effect, e , counterfactually depend on it *or* by being part of a chain of counterfactual dependencies. According to Lewis's 1973 analysis, it suffices for c to cause e that there be events, e_1, e_2, \dots, e_n such that e counterfactually depends on e_n , e_n causally depends on e_{n-1} , and so on back to e_1 , which must counterfactually depend on c . One nice feature of Lewis's analysis is that it does not have the consequence that closing Switch A did not cause the light to go on. There is an

intermediate chain of counterfactual dependencies corresponding to the current running through the wire from Switch A to the bulb. The Mackie-style approach can offer a parallel but slightly different use of chains by requiring not just that c be an NS condition for e but also that there be an event at each time between the occurrence of c and the occurrence of e such that c is an NS condition of it and it is an NS condition of e . (See Bennett 1988, p. 45.) Since Switch B's being closed is not an NS condition for any of the events that take place between the current from Switch B reaching the junction box and the light going on, such a revision would allow the Mackie-style analysis to avoid the judgment that Switch B caused the light to go on.

The problem with an appeal to the intermediate chain of events is that sometimes the preemption takes place without any time left before the effect occurs. Such cases have been compellingly described by Jonathan Schaffer (2000b).⁵ Schaffer nicely labels his cases as cases of *trumping preemption*. The major and the sergeant at exactly the same time yell to the corporal, 'Charge'. Orders from a major trump the orders from the sergeant because of his higher rank. When the corporal decides to charge, it is pretty clearly because the major ordered him to charge not because the sergeant did. But, notice the decision to charge does not counterfactually depend on the major's order—if the major hadn't hollered the corporal still would have decided to charge. Also notice that the sergeant's order is an NS condition of the charge. It is not clear that there is a chain of intermediate events that will recover only the correct causal connections for our two analyses. Here is another case of trumping preemption from Schaffer that's makes it very clear that intermediate causal chains will not be of any help:

⁵Also see Michael McDermott (1995, p. 530) and Doug Ehring (1997, pp. 21, 31).

It is a law of magic that the first spell cast on a given day match the enchantment that midnight. Suppose that at noon Merlin casts a spell (the first of the day) to turn the prince into a frog, that at 6:00 P.M. Morgana casts a spell (the only other that day) to turn the prince into a frog, and at midnight the prince becomes a frog (Schaffer 2000b, p. 161).

Lewis's appeal to an intermediate chain of events doesn't help. The sticking point is that there is no intermediate event between Merlin's spell and the enchantment—the spell acts directly. So, the enchantment doesn't counterfactually depend on any intermediate event. While the appeal to temporally intermediate events spares the Mackie-style analysis from saying that Morgana's spell caused the prince to turn into a frog, it also mistakenly rules that Merlin's spell was not a cause either. There is no event temporally between Merlin's spell and the midnight enchantment for which Merlin's spell is an NS condition.

c. Transitivity

Some relations are transitive. Identity is transitive. If $a = b$ and $b = c$, then $a = c$. More generally, relation R is transitive if and only if, for all x, y , and z , if x stands in R to y and y stands in R to z , then x stands in R to z . Is causation transitive?

Lewis assumed that causation was transitive and believed that the counterfactual dependence was not. In fact, that's the justification he gives in his 1973 paper for not identifying causation with

counterfactual dependence between actual events and instead identifying it with the ancestral of that relation. One can easily appreciate why Lewis was inclined to make that assumption. Causation is making happen. If so, how can an event make another event happen and that second event make a third event happen and it not be true that the first event also made the third event happen? Indeed, in such a case isn't it bound to be true that the first event made the third event happen by making the second event happen? At first glance anyway, transitivity seems to be an undeniable feature of the causal relation.

Here is a version of an example due to Hartry Field that at least appears to throw this common assumption for a loop. Suppose Henry places a bomb outside Joe's door and lights the fuse. Once Henry leaves, Melissa happens to arrive at Joe's place. Seeing the bomb and being a friend of Joe's, she defuses the bomb, rendering it harmless. It seems that Henry's placing the bomb in front of Joe's door caused Melissa to defuse it. It also seems that Melissa's defusing the bomb caused Joe not to be killed. But, is it true that Henry's placing the bomb outside Joe's door caused Joe not to be killed? Well, if causation is transitive, that should be true, but it certainly seems to be at least a very, very odd thing to say. What's going on?

It is interesting to contrast Lewis's counterfactual analysis with our account from Section 2 that identifies causation with counterfactual dependence. Lewis's analysis was designed with transitivity in mind and has the consequence that Henry's placing the bomb outside Joe's door did cause Joe not to be killed. (Pretty clearly, the Mackie-style account has that consequence too.) Is this a conclusion we should accept in order to preserve the transitivity of causation? Our account from Section 2 appears to have the consequence that causation is not transitive. According to it, Henry's placing the bomb outside Joe's door did not cause Joe not to be killed because it is not the case that,

if Joe hadn't put the bomb outside of Henry's door, then Joe would have been killed. No real consensus has emerged in the literature as to which of these approaches gives the more plausible result.

4. Three More Core Cases: Indeterministic Causation

All the examples in the previous two sections were assumed to take place in a deterministic world, though they also make sense without the assumption of determinism. The examples in the present section will all be assumed to take place in an indeterministic world and only make sense if they are assumed to take place in an indeterministic world. All of these examples include an assignment of a less than 100% chance to some event. Like the examples from Section 2 and Section 3, they all raise issues pertinent to understanding what causation is.

a. Chancy Causation

The first of these examples was intended to show that there can be chancy causation. Here is a presentation of such a case by Fred Dretske and Aaron Snyder:

Box R contains a randomizing device; once activated it proceeds, in a perfectly random manner, to one of its one hundred different terminal states. Each of the

terminal states may be supposed to be equally probable so that the probability of the box ending in state number 17 is 0.01. One can think of the device as embodying certain quantum mechanical processes—e.g. the emission of an electron (the momentum of which is appropriately confined by some slit) towards a screen which has one hundred different areas suitably marked off as terminal states. Attached to Box R is a loaded revolver which fires when (and only when) the terminal state happens to be number 17. We take this device and place it next to a cat, point the revolver at the cat and activate the box. Things go badly for the cat; the improbable occurs and the cat is killed (1972, pp. 69-70).

As Dretske and Snyder point out about such a case, it would be very natural to accuse them of killing the cat, of having caused the cat's death. If this is correct, then it reveals something interesting and maybe even a little surprising about causation. Despite what our earlier examples might have suggested, causation doesn't have much at all to do with regularity or constant conjunction. The case is notable not just because the cause isn't in any way sufficient for the effect. It is also notable because it didn't even make the effect very likely; it gave the effect only a .01 chance of happening.

As should be pretty obvious, if activating and placing the contraption next to the cat caused the cat to die, the NS approach looks to be in serious trouble. There would be causation without anything even resembling constant conjunction. Even the complete state of the world at the time the contraption is set next to the cat fails to be sufficient for the cat's death. So, the activation and placement of the box can't be an NS condition for the death. The counterfactual account gives a different answer. Plausibly enough, if they hadn't activated the box, if they had not placed the

contraption next to the cat, then the cat wouldn't have died. So, despite being proposed with determinism in mind, our counterfactual analysis gives the result that there is causation in the Dretske-Snyder case.

It is interesting that, when Dretske and Snyder proposed their case, they were convinced that we would be tempted to say that activation of the box *did not cause the box to end up in state 17*. They think that, in this regard, the Mackie-style approach is right in line with what we would ordinarily say. We find this observation plausible! It would be odd for a referee, say, to flip a fair coin, it land heads, and then someone claim that the ref caused the coin to land heads! That does sound false; to say he caused it to land heads suggests that the flip was fixed. Somehow when the indeterminacy is front and center in what we say, we are much more reluctant to take the phenomenon as causal. Dretske and Snyder's observation opens the door for a response to their cat-case: It would be very odd to say that activating the box didn't cause the box to end up in state 17, though it did cause the death of the cat. How could it cause the death of the cat except by having causing the box to be in state 17? Maybe the Mackie-style analysis isn't in as bad shape as we thought. Maybe placing the contraption next to the cat didn't cause the cat to die. We leave this possibility as one for you, the reader, to explore.

For better or worse, it has been the overwhelmingly popular judgment of the metaphysics community that there is chancy causation. We attribute that partly to the popularity of the counterfactual analysis and its judgment in the Dretske-Snyder case. Another part of the motivation for this judgment stems from the thought that the actual world, our universe, may be indeterministic. If the lack of a deterministic connection from the present state of the world to any future states is enough to undermine any causal connection between us and any future states of the world, then there

would be no causation of those future states. Denying the possibility of chancy causation may be tantamount to denying that there is any causation in our world. Given how central causation is to our conceptual framework, that would be tantamount to denying that there are any causal processes; so no molecular bonding, no planetary rotation, no human decisions, and no life.

No, the more common reaction to the Dretske-Snyder case is to try to accommodate probabilistic causation of this sort using an appropriate analysis. The counterfactual analysis handles the simple case but won't work on other simple cases: Suppose fair roulette wheels are genuinely indeterministic, that, even given the complete state of the world at the time the ball is released, the laws of nature do not determine whether the ball will settle on a red or a black space. Now, consider an otherwise fair roulette wheel with a hidden switch that activates a series of magnets that attracts the metal ball to Red 32. The croupier drops the ball and hits the switch and the ball eventually settles on Red 32. It seems that the croupier's flipping the switch caused the ball to land in a red slot. But, notice, if the switch had not been thrown, then the ball might still have landed on red. So, it would have been false that the ball wouldn't have landed on red. If causation was just counterfactual dependence between occurrent events, throwing the switch would not have caused the ball to land on red. Following an idea of David Lewis (1986, p. 175), a natural way to revise our counterfactual approach would be as follows:

Causation as Probabilistic Counterfactual Dependence

c caused *e* if and only if, if *c* hadn't occurred, then the chance of *e*'s occurring would been much less than it actually was.

There is another popular way of analyzing causation that allows for chancy causation. Instead of analyzing causation in terms of counterfactuals about probability, the idea is to analyze causation in terms of conditional probabilities. The basic idea is that causes raise the conditional probability of their effect, that the chance that e occurs should be greater given that c occurs than given that c doesn't occur. (See Eells 1991.) This basic idea looks promising given the two cases discussed in this subsection. Placing the contraption next to the cat certainly raises the probability that the cat will be killed (even though the probability that the cat is killed never gets above 0.01). By hitting the switch on the roulette table, the probability that the ball lands red is raised from 0.50, all the way up to 1.00.

b. Overlapping

Imagine that Merlin casts a spell with a .5 chance of turning the king and prince into frogs, that Morgana casts a spell with a (probabilistically independent) .5 chance of turning the prince and queen into frogs, and that the king and prince, but not the queen, then turn into frogs (Schaffer 2000a, p. 40).

This is labeled a case of overlapping because the effects intended by Morgana and Merlin overlap. The witch and the sorcerer are both trying to turn the prince into a frog. The overlap is partial, though. Through her single spell, Morgana wants to also turn the queen into a frog; while through his single spell, Merlin also means to turn the king in to a frog. It is assumed that, when they work, spells work directly, not through any intermediate events.

The causal facts about this case seem to be pretty straightforward (still assuming that chancy causation is possible at all). Since it was the king and the prince, not the queen and prince that became amphibians, it was Merlin's spell that was effective; Merlin, not Morgana, caused the prince to be a frog. But, these facts cut to the heart of the standard ways of dealing with probabilistic causation. If we consider the conditional probabilities, the probability of the prince turning into a frog given that Morgana casts her spell is greater than the conditional probability that the prince turns into a frog given that Morgana didn't cast her spell. Morgana's spell definitely raises the probability that the prince meets the amphibious fate. If we consider the counterfactuals, it is clear that if Morgana had not cast her spell, then the chance that the prince would become a frog would have been significantly less than it actually was. The standard ways of analyzing causation seem bound to get the case wrong; they say Morgana's spell was effective.

e. Underdetermination

Underdetermination cases take matters one step further. Suppose Merlin and Morgana both cast spells with a .5 chance of turning the prince into a frog. Neither is concerned with anyone else. They are both just after the prince. Like the previous case, this cause involves overlapping; it is just that now the overlap is complete. What happens is that the prince turns into a frog (Schaffer 2000a, p. 45).

Did Morgana turn the prince into a frog? Did Merlin? There seem to be at least two equally intuitive possibilities here that cannot be easily dismissed. The first is that Merlin did and that

Morgana did not. The second is that Morgana did and Merlin did not. Nothing about the situation seems to say which is the case. There appear to be two equally good possibilities here and nothing about the situation that decides what causes what. The key causal facts in this case seem not to be determined by the probabilities, nor by any facts about the putative causes or the effect, nor by any causal chains between them.⁶ Since both possibilities seem equally good, let's just suppose that it was Merlin who beat the 50-50 odds. His spell worked. We'll also suppose that Morgana's didn't beat those odds; it lost out. As before, it turns out that Morgana's casting her spell stands in all the right probabilistic relations for it to have been the cause. Her casting the spell raised the probability of the prince turning into a frog no matter whether we cash out probability-raising in terms of conditional probabilities or in terms of counterfactuals about chance. It looks like trouble for standard probabilistic analyses of causation.

This case of underdetermination, a case of complete overlap, is potentially a serious challenge to the *possibility* of analyzing causation. In the partial overlap case there was the fact that the king turned into the frog that made it clear that it was Merlin's spell, not Morgana's, that was effective. The presence of that fact gives some hope to those who want to analyze causation. There is at least a symptom indicating that there might be some underlying truthmaker for the causal facts. Lewis shows the following flicker of hope:

⁶We are setting aside the possibility that both spells worked. That seems fair since we could have built into the case that, in an appropriate number of cases when two spells are cast at exactly the same target, the target turns into a frog that is twice as green and twice as small as someone who is the victim of a one-spell transmutation.

We want to say that the [probability-]raising that counts is the raising of the probability of the causal chain of events and absences whereby the effect was actually caused. Raising the probability of some unactualized alternative causal chain leading to the same effect doesn't count. But it would be circular to say it that way within an analysis of causation. I hope there is some non-circular way to say much the same thing, but I have none to offer.

The complete overlap case offers no hope at all. Nothing in the example suggests that there is any fact that might serve as a truthmaker for the claim that Merlin was the cause.

There are different reactions one can have to this sort of underdetermination case. Some will object that the example is somehow faulty, that the thought that there could be a possible world where Merlin was the cause and a different possible world where Morgana was the cause plays on some sort of mistaken intuitions we have about causation. Most philosophers who have defended an analysis of causation have taken this sort of approach, including Schaffer. Some accept the example at face value, holding that causation is such a basic part of our conceptual framework that there is nothing very interesting that can be said by way of an analysis or definition of causation. Others will find solace in another idea sometimes thought to also be found in the writings of David Hume. This is the idea that causal talk is not really fact stating in the first place. These *projectivists* will say that our causal sentences project certain attitudes of ours rather than express beliefs about the way reality is; causal sentences are not true or false in the strictest and most literal sense of the words 'true' and 'false'.

6. Causal Eliminativism

The underdetermination cases are not the only arguments that make philosophers worry about the metaphysical reality of causation. Some philosophers worry that the absence of the word ‘causes’ from the formulation of fundamental theories of physics is an indication that causation is merely a folk concept, maybe like the concept of a witch, that may get lots of use in ordinary conversation, but which has no application to the world since there are no witches. Our best physical theories include fundamental laws that are equations relating various properties to other properties but without explicitly stating that there are any causal connections or even that there would be certain causal connections if certain conditions were to come to pass.

Bertrand Russell (1912-13) thought that two different features of physical theory show that our ordinary notion of causation is not actually exemplified. The first was that, going by our ordinary ascriptions of causation, it seems that causation exhibits directionality in that, in general, maybe even always, it is earlier events that cause later events, not the other way around. But, he argued, the best confirmed physical theories do not reveal any temporal asymmetry insofar as causal connections from past to future or future to past are concerned. The second is that, again going by ordinary conversations, it appears that causes are thought to be fairly localized, easily identifiable sorts of things. But, given the way physical theory is, that can’t really be true. The scratch couldn’t have caused the match to light, the match also had to be dry, not already been lit, The laws of physics are such that local information can’t determine that any effect will occur or even that an event will have a certain probability. There are too many things that didn’t go on or might have gone on that, according to the laws, were required to happen in order for an actual event to have occurred or even

to have its probability raised. If you think causes must be sufficient for their effects or even sufficient for their effects to have some positive chance of occurring, you should have big doubts about whether anything we normally refer to as a cause could really be a cause of anything. Maybe our causal talk isn't meant to describe how the world is.

The current state of unrest for work on causation may lead some to despair; the troubling examples for the analyses of causation we have considered are many and varied, and the potential conflicts with physical theory are serious. Despair is natural, but we think that is not the best reaction to have to the current state of the philosophical literature. We think this is a really exciting time for metaphysicians. Philosophers are not doing drudge work; they are not digging in their heels trying to defend their favorite theory, holding whatever convenient position is necessary to do so. Rather they are, somewhat independently of specific theories though with an eye to the insights associated with representative theories, revisiting some very fundamental issues in an open-minded and provocative manner. The questions are not what's wrong with this theory and is there any way of revising the theory to avoid the problem. In stead the questions are: Is causation transitive? What causes what in cases of overdetermination? Is there a metaphysical difference between causes and conditions? Is there something about how the verb 'to cause' works in our language that gives rise to puzzle cases? These are issues that are engaging ones, and ones that promise reward.

REFERENCES

Beauchamp, T. and Rosenberg, A. (1981) *Hume and the Problem of Causation*. New York: Oxford University Press.

Bennett, J. (1988) *Events and their Names*. Indianapolis: Hackett Publishing Company.

Dretske, F. and Snyder, A. (1972) "Causal Irregularity" *Philosophy of Science* 39, 69-71.

Eells, E. (1991) *Probabilistic Causality*. Cambridge: Cambridge University Press.

Ehring, D. (1997) *Causation and Persistence*. New York: Oxford University Press.

Hume, D. (1955) *An Inquiry Concerning Human Understanding*. Indianapolis: Bobbs-Merrill Educational Publishing.

Kim, J. (1974) "Noncausal Connections" *Noûs* 8, 41-52

Lewis, D. (1973) "Causation" *Journal of Philosophy* 70, 556-567.

Lewis, D. (1986) *Philosophical Papers, Volume II*. New York: Oxford University Press.

Mackie, J. (1974) *The Cement of the Universe*. Oxford: The Clarendon Press.

McDermott, M. (1995) "Redundant Causation" *British Journal for the Philosophy of Science* 46, 523-544.

Menzies, P. and Price, H. (1993) "Causation as Secondary Quality" *British Journal for the Philosophy of Science* 44, 187-205.

Russell, B. (1912-13) "On the Notion of Cause" *Proceedings of the Aristotelian Society* 13, 1-26.

Schaffer, J. (2000a) "Overlappings: Probability-raising without Causation" *Australasian Journal of Philosophy* 70, 40-46.

Schaffer, J. (2000b) "Trumping Preemption" *Journal of Philosophy* 97, 161-185.