2 Causation

from An Introduction to Metaphysics (2010)

by John Carroll

2.1 A familiar, central, and tricky relation

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 a biscuit from across the dinner table and bump a glass with your elbow. Water from the glass spills. Not much could be more obvious, it seems, than that you bumped the glass and thereby *caused* the water to spill. To put this in a grand-sounding way, your bump of the glass stood in the relation of *causation* to the spill. That is what we are going to try to understand better in this chapter, that relation that holds between the bump and the spill in this mundane example. This is a terrific topic because causation is about as familiar, central, and tricky as metaphysical concepts come.

The spilled-water case should be enough to convince you that causation is a familiar concept. That it is a central concept is also straightforward: it is something we do whenever we affect what is around us. It is something we undergo whenever we are affected by what is around us. Molecular bonding, planetary rotation, human decisions, and life itself are all causal processes. 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 what caused it to happen. Causation is part of philosophy too. It is so, in part, because philosophers try to improve our understanding of causation. Causation also plays a role in philosophy when the focus is on other matters. As we will see in Chapter 6, a key to the nature of mind is whether mental states can affect the material world. As we will see in Chapter 9, philosophers doing ontology wonder whether we should believe in abstract entities like Platonic universals if these entities do not participate in causal relations.

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 metaphysicians about how to understand causation better. But causation's trickiness amounts to more than just that. (Disagreement is common in philosophy.) There is something about causation that makes theorizing about it especially challenging. Unlike 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 and Determinism), or The A Theory vs. The B Theory (about time). No, the study of causation has recently come to be better defined by a range of important examples. The interesting disagreements are more about what causes what in the examples – about where the causation lies – than they are about which theory gives the correct verdict about each example. In other words, we metaphysicians have been questioning what the correct verdicts are! Though we will sketch a handful of simplified but still representative accounts of causation, the primary goal of this chapter will be to present the important examples.

2.2 The relation and the relata

Before getting to the examples, there are two preliminary matters that need to be addressed. We need to restrict our attention to certain uses of the verb 'to cause'. Something also needs to be said about what causation relates.

No doubt you have seen or at least heard the US Surgeon General's warning, "Smoking causes lung cancer." That statement is not describing any particular smoking event. It does not mention any specific individual at any specific time or place doing any causing of anything. In this way, this claim is different from the claim that the bump of the glass caused the water to spill, which, in our opening example, was very much about what you did at a specific time and place. It may be helpful to think of the warning about smoking as describing a causal relation between two properties; maybe the Surgeon General is saying that the property of smoking causes the property of having cancer. Or, perhaps the warning 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 different kind of general remark, just saying that the number of cases where smoking causes cancer is high. Whatever is being said, 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 examples of *single-case* causation, examples where both the cause and the effect are particulars – not properties.[1] The relevant causation sentences are ones that are nothing like generalizations.

Here is an assortment of relevant causation sentences that all seem plausible when taken to be about the spilled-water case:

(1) The bump caused the spill.
(2) You caused the spill.
(3) Your bumping the glass caused the spill.
(4) That you bumped the glass caused the spill.

Though these are all single-case causation sentences, they appear to refer to a variety of different kinds of causes. Sentence (1) refers to the bump, an *event*, as the cause. Sentence (2) says that you, a *person* or *agent*, caused the water to spill. Sentence (3) says that the cause was your bumping the glass; some philosophers will take that to be an event, others take it to be a *state of affairs*. In (4), the cause is that you bumped the glass, a *proposition* – more specifically, a *fact* (a true proposition). We won't bother, but we could have displayed an assortment of kinds of effects too.

The variety of causal relata has led to a remarkable number of philosophical reactions. Some metaphysicians have been inclined to think that there is more than one (single-case) causal relation. They think, for example, that *agent causation* is not *event causation* and that it is not *fact causation* either, that these three relations are distinct and call for different philosophical accounts. Others have thought that there is just one causal relation, but that it can relate many different kinds of particulars. 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! For example, some think events are unstructured individuals;[2]

[1] Chapter 9 takes a detailed look at properties and their relation to particulars.
[2] For example, Davidson, "Causal Relations" and "The Individuation of Events."

others think that all events are structured – that they all consist of an individual having a property at a time (e.g., you having the property of reaching at supper time).[3]

We will do our best not to get involved with these disputes about the causal relata. In presenting some theories of causation, we will take '*c* caused *e*' as our focus locution,[4] 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 will do this just to regiment our own discussion. It is not meant to reflect any important assumptions of our discussion. With that in mind, we will also work diligently 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 facts, whether events, states of affairs, or facts are structured or unstructured things. We will not have much to say in this chapter about sentences like (2) and the corresponding concept of *agent causation*. Agent causation will only come up again toward the end of Chapter 3.

2.3 Three theories of causation

It goes without saying that a cause causes an effect only if both the cause and the effect occur. So every theory of causation subscribes to *c* caused *e* only if *c* and *e* both take place. That causes and effects must take place in order to be causes and effects is about the only thing that is not controversial about causation. We won't explicitly put this necessary condition in our statement of the theories of causation.

[3] For example, Kim, "Events as Property Exemplifications."

[4] We will not resist using phrases of the form, 'caused *x* to *F*', which are a common and natural way of formulating causal claims. Also, we will 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 (1) somehow makes the suggestion that the bump was the *only* cause of the spill, whereas saying, "The bump was a cause of the spill" seems to make the suggestion that there were other causes as well (cf., Unger, "The Uniqueness in Causation"). We suspect that this difference has more to do with the finer points of language usage than with the nature 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?

One idea that has played a central role in the history of philosophy since David Hume is that causation is a matter of two things always being conjoined. Hume said, "We may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second."[5] 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:

Constant Conjunction
c causes[6] *e* if and only if there are properties *F* and *G*, such that *c* has *F* and *e* has *G*, and each event of kind *F* is followed by an event of kind *G*.

As illustration, consider Pompeii and Mt. Vesuvius. 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 was heading with a specific flow rate toward Pompeii. We would need to fill this in more, especially by 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, destruction of the nearby city would follow. So, according to Constant Conjunction, 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 held anything quite as simplistic as Constant Conjunction as his or her official account of causation.[7] To see one reason why it is simplistic, note that it might be true just as a matter of pure coincidence that every coin that has ever been in a

[5] Hume, *An Inquiry Concerning Human Understanding*, p. 79, first published in 1748.

[6] The use of the present tense ('causes') here to describe causation between two (particular) events is grammatically odd, because it suggests that the relation expressed is constant or repeatable in some way. It would be more natural to use the past tense ('caused'), the present progressive tense ('is causing'), or the future tense ('will cause'). Since we do intend the account to apply to any pair of events – past, present, or future – and since it would be tedious to include all the tense variations, we have chosen to say 'causes'. If your grammar sensibilities are offended, we apologise. We will follow this same convention with all of the accounts of causation to be presented.

[7] Beauchamp and Rosenberg devote their book, *Hume and the Problem of Causation*, to spelling out the modifications needed to make a Constant-Conjunction account of causation much more tenable.

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 that pocket of those pants; the pants will be destroyed in a fire tonight. So, according to Constant Conjunction, that coin being in those pants caused it to be a nickel; anytime any event involving a coin being in that pocket occurs, it is followed by an event of a coin being a nickel. Of course, that is not at all a plausible consequence of Constant Conjunction. Any theory of causation that gives us a result like this needs to be reworked or dismissed. To avoid counterexamples of this sort, proponents of the idea that causation is Constant Conjunction will insist that the constant conjunction has to be of the right sort. Not just any true regularity connecting one kind of event with another can underwrite causal truths. Traditionally, proponents of this approach have argued that only laws of nature (or certain sorts of laws of nature) have that status.

One descendant of Constant Conjunction is an account defended by J. L. Mackie and also by Jonathan Bennett.[8] Greatly simplified, the idea is that what is important to causation is that the cause be an *ns condition* of the effect, that the cause occurring be a *necessary* part of a condition that together with the laws of nature is *sufficient* for the effect to occur.

NS Condition
c causes *e* if and only if *c* is an ns condition for *e*.

Consider Mt. Vesuvius erupting and Pompeii 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 laws of nature would not entail that Pompeii would be destroyed. The idea is that the eruption was a crucial part of a certain portion of a time-slice of our 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.

[8] See Mackie, *The Cement of the Universe*; and Bennett, *Events and Their Names*.

In 1973, David Lewis made popular and sophisticated a different idea.[9] It is somewhat similar to NS Condition in that both ideas are based on the thought that the cause has to somehow be necessary for the effect. Suppose that Pedro is standing alone near an old abandoned house and whips a baseball at one of the windows. The window shatters. Evidently, the throw caused the window to shatter. But it also seems perfectly true that, if Pedro hadn't thrown the baseball, then the window wouldn't have shattered. Maybe causation between two events just amounts to it being the case that, if one event had not occurred, then the other would not have occurred.

Counterfactual Dependence
c causes *e* if and only if, if *c* weren't to occur, then *e* wouldn't occur.

Notice that the theory includes a counterfactual conditional, an 'if–then' sentence in the subjunctive mood.[10] It essentially says that there is causation between *c* and *e* if and only if *e* counterfactually depends on *c*. In a straightforward way, this theory avoids the problem offered above as trouble for Constant Conjunction. A more interesting relation is required to hold between the events than mere constant conjunction. It is not true that, if the coin weren't in those pants, then it wouldn't have been a 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. If the bump hadn't occurred, then neither would have the spill.

Many more accounts 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 to be set aside are the transference theories and the manipulability theories.[11]

[9] Lewis, "Causation."

[10] The subjunctive mood is important. Compare: 'If Oswald didn't shoot Kennedy, then someone else did', which is in the indicative mood and clearly true, with 'If Oswald hadn't shot Kennedy, then someone else would have', which is in the subjunctive mood and quite doubtful unless you have been convinced by some conspiracy theory about JFK's assassination.

[11] Recent work: Ehring defends a trope transference theory in *Causation and Persistence*. Menzies and Price defend a manipulability theory in "Causation as Secondary

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 *c* causes *e* if and only if bringing about *c* would be an effective means for bringing about *e*. The main challenge here is whether these accounts can be non-circular or reductive in such a way as to be suitably illuminating, the worry being that bringing about and causing seem to be very similar concepts.

2.4 Three core issues

2.4.1 Lay of the land

In this section, we will introduce three core issues that all bring out some interesting features of causation and also raise at least preliminary trouble for either NS Condition or Counterfactual Dependence. (Constant Conjunction will also be discussed, but to a lesser degree.) Three more core issues will be introduced in 2.5.

2.4.2 Non-causal connections

We considered one case of a non-causal connection in 2.3, that of the nickel in the new pair of pants. It presented a serious problem for Constant Conjunction because being a coin in that pocket of those pants is constantly conjoined with being a nickel, though there is not a corresponding causal connection between the coin being in that pocket and that coin being a nickel.

A different kind of non-causal connection arises with *epiphenomena*. Let's start this example by making a certain simplifying assumption: that it is true and a law of nature that, whenever a properly functioning barometer's reading drops, a storm occurs shortly thereafter.[12] Then suppose a

Quality". In *Making Things Happen*, Woodward defends a non-reductive descendent of manipulability theories. He analyzes causation in terms of a causal notion of intervention.

[12] Even with the properly functioning clause, and even assuming it is true that, whenever a properly functioning barometer's reading drops, a storm occurs shortly thereafter,

storm is approaching your hometown, the atmospheric pressure drops dramatically, and your properly functioning barometer records the drop in pressure and displays the new reading. You predict there will be a storm. Your prediction is true; your hometown is deluged. In this example, there is a constant conjunction between changes in the barometric readings and storms. That obviously presents a problem for Constant Conjunction. It will have the consequence that the change in the barometric reading, what you 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 phenomenon not involved in the processes producing the storm in your hometown. The difference between epiphenomena and causes is an important one. For example, in the study of disease, it is crucial for doctors to be able to distinguish the causes of the disease from its symptoms.

The change in the barometric 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, like Constant Conjunction, NS Condition has the unwanted consequence that the change in the reading caused the storm to arrive. For Counterfactual Dependence, everything depends on what would be the case if the reading hadn't dropped. If there had still 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, Counterfactual Dependence would give the right answer: the change in reading did not cause the storm. But is that what would happen if the barometer hadn't shown a drop in pressure? Maybe, instead, in evaluating the key conditional, we should "backtrack" concluding that, if the reading hadn't dropped, then there wouldn't have been a drop in atmospheric pressure. The barometer would have kept working perfectly; it's just that there wouldn't have been a drop

this is probably not really a law of nature – the barometer could be functioning properly and be inside a decompression chamber when the reading drops without any storm following shortly thereafter. Philosophers have never objected to the barometer example on this score. The simplicity of the alignment of the causal connections indicates that corresponding laws are possible, and the possibility of the example is all that is important as regards testing metaphysical accounts of causation.

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. For this reason, defenders of counterfactual theories of causation often insist that the key conditional not be understood in such a backtracking manner.

A third kind of non-causal connection arises because of certain strong connections between events. When Socrates drank the hemlock and died in the Athenian prison, his wife, Xanthippe, became a widow.[13] 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 dying is an ns condition of Xanthippe becoming a widow. So, Counterfactual Dependence and NS Condition both count Socrates' death as a cause of Xanthippe 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. The connection seems to be too much a conceptual matter. To drive home this point, 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.

This, our final example of a non-causal connection, is perplexing for a couple of reasons. For one, this may be a case where our simplifying approach to events becomes an issue. One might argue that Xanthippe becoming a widow is not an event in virtue of it not being suitably concrete or in virtue of it not being an intrinsic change in Xanthippe. For another reason, whether it is an event or, say, a state of affairs, it is odd to think that Xanthippe becoming a widow was *uncaused*. Yet if it were not uncaused, what else could have caused it other than Socrates' death? Can you think of a better candidate?

2.4.3 Simultaneous causation

There is an important part of Constant Conjunction that we have not said much about. It requires that the events of kind *F* are always *followed by* an

[13] Kim, "Noncausal Connections," pp. 41–2.

event of kind *G*. You were probably just assuming that causes always occur before their effects, and that this was an implicit part of NS Condition and Counterfactual Dependence too. Well, it is 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 Constant Conjunction.

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. Suppose also that you push down on one side and the other side goes up. Then the one side 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 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 accounts set out in 2.3. For instance, consider Counterfactual Dependence. If the side you pushed hadn't gone down at the time it did, then the other side would not have gone up at exactly that same time. Regarding NS Condition, your pushing down on your end of the seesaw is an ns condition of the other side rising. Thus, by the lights of both Counterfactual Dependence and NS Condition, the causation is there.

This is not to say that Counterfactual Dependence and NS Condition are not threatened by this case of simultaneous causation. Though they do have the intuitive consequence that the side you pushed down on simultaneously caused the other side to rise, they say that the other side going up simultaneously caused your side to go down! That is the trouble for these two theories. They say that the seesaw case is a case of *mutual causation*. That doesn't seem right; in this case, *c* caused *e* but *e* didn't cause *c*.

2.4.4 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 raindrops to worry about, not a bit of wind. Not surprisingly, the match lights. This story of Nelson and his match might seem boring. Indeed,

it is; any ordinary case of causation would have sufficed to raise the present issue.

The issue is whether there is any important metaphysical difference between what are causes of an effect and what are sometimes thought of as mere background conditions. Most everyone will agree that Nelson striking the match caused it to light. Most everyone will also agree that it would be at least odd in typical conversations about the cookout to say, "The presence of oxygen caused the match to light." The tough question is: Did the presence of oxygen cause the match to light or was it only a background condition?

NS Condition and Counterfactual Dependence both imply that the presence of oxygen was a cause. That does seem at least a little worrisome; it really is odd to say that the presence of oxygen caused the match to light. But, fortunately for these theories, there is a plausible explanation of why it would be odd to say this even if the presence of oxygen was a cause. The explanation is 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 it is usually odd to report what is already accepted as true. Yes, the presence of oxygen is an ns condition for the match lighting, and it is certainly true that, if oxygen hadn't been present, then the match wouldn't have lit. But it is just not so clear that the oxygen didn't cause the match to light.

The matter of background conditions gets more interesting when one realizes that these theories also count other things as causes that it would be odd even to describe as background conditions. (We will focus on Counterfactual Dependence, but all the same points could be made about NS Condition.) 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 a Martian ray gun, the match would not have lit. If you had swiped the match from Nelson, it would not have lit. So, according to Counterfactual Dependence, 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 match. Was all that taken for granted in the conversation? If not, what explains the oddness of reporting these remarkable consequences of Counterfactual Dependence? It might be that these consequences of Counterfactual Dependence are false.

2.5 Three more core issues

2.5.1 You are here

Mostly to make sure that 2.4 isn't too long, we have saved three core matters for this fifth section. The cases to be discussed here also have a lot in common with each other, including that they all get a lot of attention in the recent metaphysical literature on causation.

2.5.2 Overdetermination

The standard story here is of a deserter sentenced to death. To keep it simple, we will suppose that the firing squad includes just two members, A and B, each of whom is a crack shot. The commanding officer gives the order to fire and the sharpshooters shoot their loaded rifles at exactly the same time. Side by side, the bullets from the two rifles simultaneously pierce 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 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 remaining causal truths are about this situation. We need to consider whether A's shot caused the deserter's death. We also need to consider whether B's shot did.

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. It would be strange if somehow the shots lacked that effect in virtue of the fact that they both occur. (Notice there is no significant interaction between the two shots.) These considerations, if telling, would stir up a whole lot of trouble for Counterfactual Dependence. 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 Counterfactual Dependence, A's shot didn't cause the death. The same goes for B's shot. These consequences of Counterfactual Dependence contradict the tempting conclusion that each shot was a cause of the death.

But hold on. Don't dismiss Counterfactual Dependence too quickly! This is not a case where our intuitions are very clear or very strong. Maybe this theory has things right. Notice that A's shot didn't really make any difference regarding the deserter's death; the deserter would have died even if A had felt a last-second twinge of remorse and hadn't shot. The same point applies to B's shot. So 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 the causation in this case, then Counterfactual Dependence seems to get things exactly right. NS Condition will agree that at least one shot being fired caused the death, but, as we have seen, it also holds the now not-so-obvious consequence that it is true that A's shot caused the death and that B's shot did as well. NS Condition and Counterfactual Dependence make conflicting judgments about our overdetermination case, and yet both judgments have a certain amount of plausibility.

2.5.3 Preemption

Let's consider a case where there is more agreement among metaphysicians about the causal facts. Two 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 the junction box. The 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. If both switches are closed simultaneously and so the current from both switches reaches the junction box at the same time, 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 box, 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?

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 just Switch A or just Switch B hadn't been closed, the light bulb still would have gone on, just as if either just Sharpshooter A or just Sharpshooter B hadn't fired, then the deserter still would have died. So, as in the overdetermination case, neither of our two potential causes seems to make a difference with respect to the effect. In other ways, preemption cases are unlike 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. It is for this reason that the widely accepted judgment on this preemption case is that Switch A caused the light bulb to go on and that Switch B did not.

Preemption cases have been a thorn in the side of a wide range of theories about causation. Counterfactual Dependence has the unintuitive consequence that closing Switch A did not cause the light bulb to go on; if Switch A hadn't been closed, then the light still would have gone on. The problem for NS Condition is different. It correctly counts that Switch A was closed as a cause of the light going on; the problem is that it also counts that Switch B was closed 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 a theory of causation to try to sidestep problems with preemption is to make use of some intermediate chain of events running between the cause and the effect. Lewis's counterfactual account identified causation with the *ancestral* of Counterfactual Dependence. 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, it suffices for *c* to cause *e* that there be events, e_1, e_2, …, 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 this account 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. NS Condition can be revised in a slightly different fashion

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*.[14] Since Switch B being closed is not an ns condition for any continuous event-chain that takes place between the current from Switch B reaching the junction box and the light going on, such a revision might allow NS Condition to avoid the implausible judgment that Switch B caused the light to go on.

The problem with any appeal to the intermediate chain of events is that the preempting could take place without any time left before the effect occurs. Such cases have been compellingly described by Jonathan Schaffer. Schaffer nicely labels these cases as cases of *trumping preemption*. Suppose a major and a sergeant at exactly the same time yell to the corporal, "Charge!" Orders from a major trump the orders from the sergeant because of the 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 that 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. And it is at least not clear that there is a chain of intermediate events that will recover the correct judgments for either NS Condition or Counterfactual Dependence.

Here is another case of trumping preemption that's even more compelling. It shows that intermediate causal chains will not be of any help:

> Imagine that 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 6:00pm. Morgana casts a spell (the only other that day) to turn the prince into a frog, and that at midnight the prince becomes a frog.[15]

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 NS Condition from saying that

[14] Bennett, *Events and Their Names*, p. 45.

[15] Schaffer, "Trumping Preemption," p. 165.

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.

2.5.4 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 held that counterfactual dependence was not. In fact, that's the justification he gives for not identifying causation with counterfactual dependence and instead identifying it with the ancestral of that relation. One can easily appreciate why Lewis took causation to be transitive: 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.[16] 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 placing the bomb in front of Joe's door caused Melissa to defuse it. It also seems that Melissa defusing the bomb caused Joe not to be killed. But is it true that Henry 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.

It is interesting to contrast Lewis's approach with Counterfactual Dependence. Lewis invoked chains of counterfactual dependence with

[16] Reported in Hall, "Causation and the Price of Transitivity," p. 183, and in Maslen, "Causes, Contrasts, and the Nontransitivity of Causation," p. 350.

transitivity in mind and so his approach has the consequence that Henry placing the bomb outside Joe's door did cause Joe not to be killed. Is this a conclusion we should accept in order to preserve the transitivity of causation? Hard to say. In contrast, Counterfactual Dependence appears to have the consequence that causation is not transitive. According to this theory, Henry 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 consensus has emerged in the literature as to whether we should accept or deny the transitivity of causation.

2.6 Chancy causation

All the examples in the previous two sections could have taken place in a deterministic universe, a world where the state of the world at one time, together with the laws of nature, determine the state of the world at all other times. (See Chapter 3 for a more detailed discussion of Determinism.) The examples to be introduced in the present section will all be assumed to take place in an indeterministic world. All of these examples include an assignment of a less-than-unit chance to some event. Like the examples from 2.3 and 2.4, they all raise issues pertinent to understanding what causation is.

2.6.1 Is chancy causation possible?

The first of these examples was given by Fred Dretske and Aaron Snyder in order to show that there can be chancy causation:

> 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.[17]

As Dretske and Snyder go on to say about their case, it would be very natural to accuse them of killing the cat, of having caused the cat's death.[18] 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 constant conjunction. The case is notable not just because there is no corresponding regularity and the cause isn't in any way sufficient for the effect. It is also notable because the cause didn't even make the effect likely; activating the box made the effect have a 0.01 chance of happening.

As should be pretty obvious, if activating the box caused the cat to die, NS Condition 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 of the death.

Counterfactual Dependence gives a different answer. Plausibly enough, if they had not activated the device, then the cat wouldn't have died. So despite being proposed with Determinism in mind, Counterfactual Dependence gives the result that there is causation in the Dretske–Snyder case. But Counterfactual Dependence gives questionable verdicts on other simple probabilistic 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 where it is supposed to. It seems that the croupier flipping the switch caused the ball to land in a red slot. But, notice, if the switch had not been flipped, then the ball might still have landed on red. So it would have been false that the ball wouldn't have

[17] Dretske and Snyder, "Causal Irregularity," pp. 69–70.

[18] We are assuming as Dretske and Snyder do that, in this thankfully hypothetical example, they are the miscreants who aim the revolver at the hapless cat.

landed on red. Counterfactual Dependence says that the croupier throwing the switch did not cause the ball to land on red.

Employing an idea of Lewis's,[19] a natural way to revise Counterfactual Dependence to avoid this consequence would be as follows:

> *Probabilistic Counterfactual Dependence*
> *c* causes *e* if and only if, if *c* weren't to occur, then the chance of *e*'s occurring would be much less than it actually is.

On this theory, throwing the switch caused the ball to land on red, because, if the switch hadn't been thrown, the chance of the ball landing on red would have been much less than it actually was. An element of vagueness makes it difficult to know what this view says about the cat. If Dretske and Snyder's device hadn't been activated near the cat, the chance of the cat's death certainly would have been less than it actually was; it would have been something less than 0.01. But, even if it had been zero, it is not clear whether that is *much* less than it actually was. Is zero much less than 0.01?

There is another popular way of characterizing causation that allows for chancy causation. Instead of explaining causation in terms of counterfactuals about chance, the idea is to describe 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.[20]

> *Probability Raising*
> *c* causes *e* if and only if the chance that *e* occurs given that *c* occurs is greater than the chance that *e* occurs given that *c* doesn't occur.

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 on red is raised from 0.50 all the way up to 1.00.

It is interesting that when Dretske and Snyder proposed their case, they were convinced that, though activating the device caused *the death of the cat*,

[19] See Postscript B to "Causation," in Lewis, *Philosophical Papers*, vol. II, pp. 175–184.

[20] Eells develops a sophisticated version of this approach in his *Probabilistic Causality*.

activating the device did not cause *the device to end up in state 17*. We can see why. It would be odd for a referee, say, to flip a fair coin, it land on heads, and then someone to claim that the ref caused the coin to land on heads! That does sound false; to say he caused it to land on heads suggests that the flip was fixed, that flipping the coin made it land on heads. Somehow when the indeterminacy is front and center in what we say, we are much more reluctant to take the chancy phenomena as caused. This observation opens the door for a response to Dretske and Snyder's case. It would be very odd to say that activating the device didn't cause it 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 caused the device to be in state 17? Maybe we should have concluded that activating the box didn't cause the cat to die.

We leave this possibility as one for you to explore. For better or worse, it has been the popular judgment of the metaphysics community that there is chancy causation. We attribute that partly to the popularity of counterfactual approaches and their 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 with any future states of the world, then 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. That would mean that we would be in the absurd position of having to deny that there is any molecular bonding, planetary rotation, human decision and life. Just so, the more common reaction to the Dretske–Snyder case is to acknowledge the possibility of probabilistic causation.

2.6.2 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.[21]

[21] Schaffer, "Overlappings: Probability-Raising without Causation," 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 into a frog. It is assumed that, when they work, spells work directly, not through any intermediate events.

The causal facts about this case are pretty straightforward. 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 chancy 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. Probability Raising and Probabilistic Counterfactual Dependence seem bound to get the case wrong; they say Morgana's spell was causally effective.

2.6.3 Underdetermination

Underdetermination cases take matters one step further. Suppose Merlin and Morgana both cast spells with a 0.5 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 example involves overlapping; it is just that now the overlap is complete. What happens is that the prince turns into a frog.[22]

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. The key causal facts in this case seem not

[22] Schaffer, "Overlappings: Probability-Raising without Causation," p. 45.

to be determined by the probabilities, nor by any facts about the putative causes or the effect, nor by any causal chains between them.[23] Since both possibilities seem equally good, for the moment, let's just suppose that it was Merlin who beat the 50:50 odds; it was his spell that worked. We'll also suppose that Morgana's didn't beat those odds; it lost out. As before, it turns out that Morgana casting her spell stands in probabilistic relations that often accompany a causal connection: her casting the spell raised the conditional probability of the prince turning into a frog, and, if she hadn't cast her spell, the probability of the prince turning into a frog would have been much less. Probability Raising and Probabilistic Counterfactual Dependence don't permit the result that only Merlin's spell worked, even though that seems to be a genuine possibility.

This case of underdetermination, a case of complete overlap, is potentially a serious challenge to the *possibility* of an account of 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 provide an account of 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:

> We want to say that the 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.[24]

The complete overlap case appears to dash that hope. Nothing in the example suggests that there is any fact that might serve as a truthmaker for the claim that Merlin was the cause.

[23] We 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 transmogrification.

[24] David Lewis, "Causation as Influence," pp. 79–80.

There are different reactions one can have to this sort of underdetermination case. Most philosophers who have defended one of the standard accounts of causation will object that the example is somehow faulty. They hold 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. Schaffer agrees.[25] Others accept the example at face value, holding that causation is such a basic part of our conceptual framework that there is nothing interesting that can be said by way of a metaphysical account of causation. They take (single-case) causal facts to be fundamental facts. They believe in brute causation.[26] Despite how central causation is to our conceptual framework, others will resort to the idea that there really isn't any causation. In the underdetermination case, it seems as if there is nothing at all in the world that could make it true that only Merlin's spell was a cause or that only Morgana's spell was a cause. Some such *Anti-Realists*, the *Eliminativists* (e.g., Bertrand Russell in "On the Notion of Cause"), hold that causation sentences don't succeed in describing the world, and so also hold that, strictly speaking, nothing causes anything else. Others, the *Projectivists* (e.g., Simon Blackburn in "Hume and Thick Connexions" or Huw Price in "Causal Perspectivalism") will utter sentences like "The bump caused the spill," but are Anti-Realists in virtue of thinking that such utterances will project something about us rather than convey information about the way reality is independent of us.[27]

2.7 A concluding observation

The challenging examples for the assorted accounts of causation are many and varied. It would be easy to despair about the prospects for making any progress on the topic of causation, but don't. We think this is a really

[25] Schaffer, "Causation and Laws of Nature: Reductionism."

[26] See Carroll, "Anti-Reductionism," for the history of – and the motivation for – this anti-reductive stance on causation.

[27] The underdetermination cases are not the only arguments that make philosophers worry about the reality of causation. Some 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

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, revisiting some fundamental issues in an open-minded and provocative manner. The questions are not: What's wrong with this theory? Is there any way of revising the theory to avoid the problem? Instead, the questions are: Is causation transitive? What causes what in cases of overdetermination? Is there a metaphysical difference between causes and background conditions? These are compelling issues; that they are being tackled augurs philosophical progress.

> 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.