

Causation

1. The Counterfactual Account of Causation: We've discussed regularities in nature, but what about particular instances of causation? For instance, consider Anne throwing a rock at a window followed by the window's shattering. We need an account of what it MEANS to say that the rock CAUSED the window to shatter.

Hume will of course say that we infer causation because we observe a regular pattern in nature of events of the rock-throwing sort followed by events of the window-shattering sort. The Humean view:

Constant Conjunction C caused E if and only if events of type-C are always followed by events of type-E.

But, of course, a problem with this "constant conjunction" account of causation is that we ALSO regularly observe things like alarm-clock-ringing followed by coffee-shop-opening (as in the Causing Coffee case). Hume's account entails that my alarm clock caused the coffee shop to open—which seems false. What other views of causation are there? The most popular competitor comes to us from David Lewis:

Counterfactual Dependence C caused E if and only if, had C not occurred, E would not have occurred either.

On the counterfactual account of causation, e.g., the rock-striking causes the shattering because, had the rock-striking not occurred, the window would not have shattered. This account seems initially promising. It doesn't mistakenly attribute causation in cases like:

- Every time my alarm clock rings, the shop down the street opens.
- Every time I forget to bring my umbrella, it rains.
- Every time I wear my special jersey, my team wins.

For, if the supposed cause had NOT occurred (if my alarm clock DIDN'T ring, or I DID bring my umbrella, or I DIDN'T wear my jersey), the effect still would have occurred.

2. Analyzing Counterfactual Conditionals: The counterfactual account faces many difficulties. Let's start with just one.

Common Cause: Consider the following case:

Barometer Every time the reading on the barometer drops, a storm follows.

Note that the Constant Conjunction account of causation entails that the barometer reading CAUSES the storm—which seems false. Does the Counterfactual Dependence account fare any better. **If the barometer reading hadn't dropped**, then... Then what?

*****DIGRESSION*****

The Possible Worlds Analysis of Counterfactuals: How do we assess 'counterfactual conditionals' (i.e., "if-then" claims about what would've happened, had some particular thing been different; e.g., "If I had brought my umbrella, then it would not have rained")? Answer: Typically, they are analyzed in terms of possible worlds. In order to determine whether this statement is true, we consider the possible worlds where the antecedent (I brought my umbrella) is true, and ask what happens in those worlds. But, that's not quite right. For, there's a possible world where I bring my umbrella and then cheese whiz starts flowing out of my pockets while blue whales begin falling from the sky. No. Rather, we ask what happens in those worlds that are "*nearest*" to ours (i.e., most similar to ours in some important respect).

Intuitively, in the "*nearest*" worlds where I bring my umbrella, it still rains. Therefore, the counterfactual claim above is false.

But, wait. What criterion for "*nearness*" did we use just now? What does it mean for one possible world to be "*more similar*" or "*nearer*" to ours than another? You might think that nearness is simply a function of *spatiotemporal overlap*; i.e., the fewer states of affairs a world shares in common with the actual world, the further that world is from actuality. However, in light of certain counterexamples, it is more commonly thought that spatiotemporal overlap only determines modal proximity *within that subset of worlds where actual-world laws are held fixed*. To understand why, consider this case:

Nixon and the Button Consider the following three possible worlds:

- w^* – In the actual world, Richard Nixon never pushes "the button" which would initiate a global nuclear war. Furthermore, no such war ever occurs.
- w_2 – In world w_2 , Nixon pushes the button. A global nuclear war occurs.
- w_3 – In world w_3 , Nixon also pushes the button. But, a global nuclear war never occurs because the button malfunctions.

If we were to ask, 'What would have happened if Nixon had pushed the button?', the intuitive answer is: 'A nuclear war would have occurred.' However, if counterfactuals are analyzed in terms of the nearness of possible worlds, and the nearness of a possible world is determined solely by its spatiotemporal overlap with the actual world, then the answer is rather: 'The button would have malfunctioned, and a nuclear war would not have occurred.' For, in terms of differences of states of affairs, the world where the

button malfunctions (w_3) is *much* nearer to the actual world than the world where a nuclear war occurs (w_2). After all, in w_2 , our planet became an uninhabitable war-torn wasteland over four decades ago.

In order to obtain the (seemingly correct) verdict that it is really w_2 (rather than w_3) that is nearer to the actual world, we must give priority to *laws* rather than overlap. The malfunction of the button would seem to require a tiny *miracle*, or violation of the laws. Only if the laws are given priority over spatiotemporal overlap when ordering the closeness of possible worlds are we able to preserve the intuition that w_2 is nearer to the actual world than w_3 .

*****END DIGRESSION*****

Okay, back to our question: **If the barometer reading hadn't dropped, then...** what?

In light of our digression, we might think that the answer is: **...then the storm would NOT have occurred.** After all, if we hold the laws fixed, and the barometer reading doesn't drop, then this must be because the atmospheric PRESSURE doesn't drop. Thus, no storm! The unfortunate outcome is that the counterfactual account seems to entail therefore, that the barometer reading caused the storm! (which is clearly false) Aargh!

What has gone wrong? Typically, the response is that we must not permit "backtracking" when assessing counterfactual conditionals. Our mistake was to consider the removal of the drop in the barometer reading, and then allow that this removal entail an alteration of PRIOR events (e.g., the removal of the drop in pressure itself). If we allow THAT to be the case, then the following statement comes out true:

If the barometer reading hadn't dropped, then the pressure wouldn't have dropped.

Now the change of the barometer reading counts as a cause of the change in pressure; clearly the wrong result! It would be like saying, "Well, holding the laws fixed, if the window didn't shatter, it must be that the ball was never thrown in the first place." But, then, it would be true that:

If the window hadn't shattered, then the ball wouldn't have been thrown.

Now the counterfactual account entails that the window's shattering caused the ball to be thrown! NO!! The solution is that we must hold not only the laws fixed, but the past as well (or, at least, we must change as little about the past as possible). For instance, we imagine that the barometer's reading did not drop, but the pressure DID drop. NOW we get the intuitively correct answer: **...then the storm would still have occurred.**

3. Difficult Issues: Now for some other difficult issues.

Action At a Distance: Consider:

The Widowing of Xanthippe Socrates drinks the hemlock and dies. At that moment, his wife Xanthippe becomes a widow.

The Counterfactual (C.F.) account entails that Socrates' death causes Xanthippe to become a widow. For, had he not died, she would not have become a widow. Perhaps that's right. (*What is your intuition?*) However, note that this entails that there is action at a distance. For, imagine that Xanthippe is a thousand miles away when Socrates dies. The very instant he dies, she becomes a widow. This means that the cause (his death) simultaneously produced an effect (the widowing) a thousand miles away! Thus the relation between cause and effect can break the speed-of-light barrier. (Physicists are surely made uncomfortable by this, no?) Yet, the alternative is to say that Xanthippe's widowing had no cause (absurd) or some other cause (what on Earth could that be?).

Simultaneous Causation: Consider:

See Saw You push down on one side of a perfectly rigid see-saw. At the same moment, the other side goes up.

Bowling Ball A memory-foam mattress gains a concave indentation at the same moment that a bowling ball is placed upon it.

Obviously the Humean account has made the mistake of claiming that causation only occurs when one thing is FOLLOWED BY another—for here, the cause and the effect are simultaneous.

The CF account does count the bowling ball as a cause of the indentation: If you hadn't pushed on the see-saw, the other side would not have gone up; if the bowling ball had not been placed, the indentation would not have occurred. But, it ALSO seems to count the indentation as a cause of the bowling ball being placed! And the going up of the see-saw as a cause of your side's going down (*even if we prohibit backtracking*)! For, if the concavity had not been created, the ball would not have descended into that space. If the other side of the see-saw had not gone up, your side would not have gone down.

Causes vs. Background Conditions: Consider:

Lit Match You strike a match and the match catches fire.

It is true that, had you not struck the match, it would not have caught fire. So, the striking caused the flame. But, it is ALSO true that, had there been no Oxygen present, the match would not have caught fire. So, according to the CF account, apparently the presence of Oxygen was ALSO a cause of the fire. That's counter-intuitive.

It gets worse: If your parents had not conceived a child, the match ALSO would not have lit. Nor would it have lit if the Earth never formed; or if the Big Bang hadn't occurred. Etc. Apparently ALL of these things were causes of the tiny flame of the matchstick.

[Note: David Lewis was actually happy to accept this. Strictly speaking, these things ARE all causes of the fire. However, context dictates that we don't care about them. We just sort of PRE-SUPPOSE these things as "background conditions". OTHER contexts DO pick out things like the presence of oxygen as relevant. For instance, imagine that some scientists are conducting an experiment in a vacuum. Some oxygen accidentally gets into their experiment area and it catches fire. Here, the "background condition" seems to involve a LACK of oxygen—so the context picks out the presence of oxygen as a cause of the fire.]

Absences: Note that the CF account counts absences, or failures, as causes. For instance, imagine that the lit match then sets aflame some nearby oily rags. A roaring blaze breaks out. Meanwhile, the sprinkler system above fails to go off. The house burns down. It is true that, had the sprinkler system NOT failed, the house would not have burned down. So, apparently the failure of the sprinkler system is a cause of the fire (according to the CF account). And perhaps that's right. After all, the fire inspectors WOULD probably cite such a failure in their report. However, imagine further that the town where the fire occurs sits below a large dam, which holds back a lake. It is also true that, had the dam broken, the fire would not have occurred (because the town would have been inundated with water). So apparently the dam's failure to break is ALSO a cause of the fire! *(Again, we'll have to relegate the dam's failure to break to the "background conditions" in this context. But, is that a satisfying reply?)*

Overdetermination: Consider:

Broken Window 1 Anne and Brett both simultaneously throw rocks at a window. They arrive at the same time and the window shatters.

If Anne hadn't thrown her rock, the window still would have shattered (because of Brett's rock). And if Brett hadn't thrown his rock... the window still would have shattered (because of Anne's rock). So, on the CF account, apparently neither rock-throwing is a cause of the window's shattering! Though, note that, if NEITHER of them had thrown a rock, the window would NOT have shattered. So, apparently, the COLLECTION of the two rock-throwings WAS a cause of the window's shattering. That seems weird.

Preemption: The following case is like the last one, but even worse:

Broken Window 2 Anne throws a rock at a window. A second later, Brett also throws a rock at the window. Anne's rock hits the window, and it shatters. A second later, Brett's rock flies through the newly created, shattered opening.

Again, if Anne hadn't thrown her rock, the window still would have shattered (because of Brett's rock). And similarly for Brett. So, once again, apparently neither rock-throwing was a cause of the window's shattering (though the collection of the two rock-throwings IS a cause). However, this verdict seems even MORE counter-intuitive than the previous. For, it puts the two rock-throwings on an equal footing, when it is quite clear that Anne's rock DID cause the window to break, while Brett's DID NOT. It mistakenly gives her rock no priority over Brett's when assigning causation.

*[One way out: We COULD say that events are "fine-grained", such that, had Anne not thrown her rock, that particular window-shattering would NOT have occurred. Rather, some other distinct event—a DIFFERENT window-shattering a second later—would have occurred instead. That would entail that Anne's rock DID cause the window to shatter. However, the implications are a bit strange. For instance, if you delay the start of a wedding by a minute, apparently it is a different wedding. If a villain plants a ticking time-bomb, and in my attempt to diffuse it, all I manage to do is delay the explosion by 1 second, apparently now the villain has not caused the explosion. Rather, *I* have! That seems wrong.]*

Furthermore, this reply can't handle simultaneous (or 'trumping') preemption cases, such as when a major and a sergeant simultaneously shout 'Charge!' to a corporal, who then begins charging. Apparently, neither the major's nor the sergeant's orders caused the corporal to charge—though it seems that it should be the major's orders (which outrank the sergeant's) which caused this. Furthermore, either way, the corporal's charging would have occurred at exactly the same time and so would have been numerically one and the same event.]

Transitivity: Imagine that I push a bowling ball (A) into a baseball (B), which then goes on to collide with a ping pong ball (C). A causes B, and B causes C. So, it makes sense to say that A causes C. That is, my pushing of the bowling ball was a cause of the ping pong ball's motion. Thus, intuitively, the relation between cause and effect is **transitive**. But, consider:

Bomb Diffuser Anne sets a ticking time-bomb to kill Chad. Brett sees this, and sneaks over to diffuse the bomb. Chad's life is saved.

On the CF account, if Anne had not set the bomb, Brett would not have diffused it. So, her setting of the bomb was a cause of his diffusing it. Furthermore, if Brett had not diffused it, I would not have lived. So, his diffusing of the bomb was a cause of my continuing to live. By transitivity, then, Anne's setting of the bomb was a cause of my continuing to live. Wait... That doesn't seem right... Perhaps causation is not transitive after all? But, that seems weird too.

4. Conclusion: As we have seen, the nature of causation is still quite poorly understood. Even our best account of what causation IS is fraught with difficulties. What's worse, as we'll see next time, this spells trouble for ethics. For, one of the foremost principles of ethics is, **Do no harm**, or in other words, Do not cause harm to others.