

# Possible Worlds and Counterfactuals: Critique and Commentary on Complicating Causation

Roman Feiman

Philosophy, particularly the discipline of metaphysics, has long concerned itself with the problem of causation. Though the nature of causation may seem intuitively obvious or axiomatic, it is precisely due to these impressions that it deserves deep philosophical investigation. There is never any doubt that causation exists in some way (that is, that events somehow 'lead' to other events). However, what exactly is meant by saying "A causes B" is not so clear. The difficulty of this question becomes more evident when one looks at the long string of illustrious philosophers who have attempted its solution – from David Hume's anti-realist regularity theory of causation,<sup>1</sup> to John Mackie's more elaborate explanation<sup>2</sup> in a similar tradition. The most recent of these propositions, and one that has proven particularly influential, is David Lewis's theory of possible worlds.<sup>3</sup>

This paper will analyze Lewis's view, examining both its advantages and its flaws. It will examine the viability of the entire "possible worlds" enterprise, looking at two traditional critiques of this approach as well as a more original argument. Of particular interest will be the phenomenon of preemption and the problems this phenomenon presents for Lewis's view. Though pre-

---

Roman Feiman will graduate from the University of Toronto this spring, with concentrations in psychology, cognitive science, and philosophy. His philosophical interests are mainly in the philosophy of mind and its relation to psychology; specifically, the nature, structure, origin, and acquisition of concepts. More recently, he has become interested in the interactions between metaphysics, epistemology and the philosophy of mind, and how these apply to processes of detecting and creating patterns, abstractions, and concepts out of an infinitely "messy" world. Roman intends to pursue this line of work through graduate work in psychology and cognitive science.

---

---

emption has long been problematic for many theories of causation, it will be argued here that the issue is in fact much less serious than it appears, not only for Lewis's account, but for all others as well. Finally, the viability of Lewis's theory will be examined in light of both its flaws and advantages.

In David Lewis's account, a statement of causation is similar to a counterfactual statement.<sup>4</sup> Saying "A caused B" is thus much like saying "If A had occurred, B would have occurred also, and if A had not occurred, then B would not have occurred." It is important to note the difference between counterfactual conditional statements and material conditional statements, which are usually of the form "If A, then B." In a material conditional, B is true only if A is, and to check whether A is true one need only compare its claim to the state of the world. Counterfactuals have a different nature – rather than making claims about what *is*, counterfactuals concern themselves with what *could be*. So if A were true, what would be the consequences for B? Here, a problem quickly arises. How can one check the validity of any given counterfactual conditional? A brief illustration: while the truth of the material conditional, "If Jimmy drove, he crashed his car" is easy to check through reference to real events, how would one check the counterfactual, "If Jimmy had driven, he would have crashed his car?" Since this statement does not concern what actually happened, its validity cannot be determined by reference to any real event. Rather, information might be required on how safe Jimmy's driving is generally, what car he drives, the road conditions and, in fact, an infinite number of other factors.

It seems, therefore, that any theory that uses counterfactuals requires some scheme by which their validity can be determined. Lewis was aware of this issue and proposed a solution that serves as the basis for his theory – the idea of possible worlds. The claim is that our universe is only one of an infinite number of possible worlds, and that each of these worlds is different in some respect; anything from a minor detail such as the color of a given chair, to things as drastic as the laws of physics. In fact, the only laws that all possible worlds must follow are the laws of

---

logic, such as the law of non-contradiction, which says that no proposition can be both true and false. Using this idea, Lewis claims that in order to check whether a counterfactual conditional is true, one must simply look at the nearest possible world in which the prerequisite condition (called “A” above) holds. The nearest possible world is one in which as many aspects of the world as possible are similar to ours. So if A is true in our world, then ours is the nearest possible world required. If A does not hold, then by definition the possible world of interest is the one where absolutely everything is identical to our world, except that A holds. For instance, if a counterfactual is concerned with a specific red chair being blue, then the nearest possible world that must be examined is the one where that chair indeed happens to be blue, but everything else is unchanged from our world. Lewis proposes that in this nearest possible world it is possible to turn the original counterfactual claim of interest into a material conditional claim; that is, since A is true in this world of interest, we should check if B is also true. If that is the case, then the material conditional holds in this possible world, which logically implies that the original counterfactual holds in ours.

Causation is, in fact, a bit more complicated. To claim that event A caused event B, one must first validate the corresponding counterfactual (what Lewis calls “causal dependence”), but there also must be a chain of events of any length, termed a “causal chain,” leading from event A to event B, with every event causally dependant on its predecessor. With this scheme, it is worth noting that while causal dependence is sufficient for causation (by definition, it is simply a causal chain with two members), causation is not simply causal dependence. This is because causation is a transitive relationship (A can cause B, can cause C, etc.) while causal dependence is not.

While this approach may seem promising, it is not without its flaws. The most obvious argument that can be leveled at the possible worlds theory is that it is too demanding, metaphysically. This argument is something of a *reductio ad absurdum* – the kind of use of possible worlds that Lewis proposes implies that they exist, a notion that is absurd and impossible. After all, if one ac-

---

---

cepts the existence of infinite possible worlds somewhere in the realm of reality, one might as well accept ghosts, witches or any other number of potentially convenient explanatory agents that cannot be shown to exist. This argument is even more cogent given Lewis's conviction that possible worlds are in fact real entities,<sup>5</sup> to the point that we are only one of these worlds, such that a person in another possible world considering us is no less real than we are while considering that person's possible world.

However, if one rejects this belief, there does not appear to be any reason to be convinced by the above criticism. Though it may seem devastating at first to Lewis's theory to reject the idea that possible worlds are real entities in the physical non-abstract sense in which Lewis views them, this is not the case. After all, convenient logical devices are commonly used without any need to establish a physical existence. One obvious example comes from our use of numbers. Though whether abstract objects such as numbers are in some sense "real" is still an ongoing debate, there is no doubt that numbers are not physical entities in the sense that possible worlds are claimed to be. Nevertheless, and regardless of whether they are real or not, it is difficult to deny their theoretical usefulness. In a similar way, there is no need to claim that possible worlds exist as physical entities in order to employ them for testing counterfactuals. Granting possible worlds the same ontological status as abstract objects seems more than sufficient.

Another issue that has long been used to challenge most theories of causation is the question of preemption.<sup>6</sup> Preemption occurs when some event A causes event B, but if A had not caused event B, then some event A<sub>-2</sub> would have caused B. The issue for Lewis's approach arises from his use of counterfactuals in the concept of causal dependence. If B would have occurred regardless of A, then it would seem that A did not cause B because one of the two conditions of Lewis's causal dependence does not hold. If A had not occurred, then B still would have occurred.

However, there is a problem with the idea of preemption, and with the criticism it implies: is the B that occurs due to A the

same as that B that would have occurred due to A-2? An example will illustrate the issue. If Jimmy crashes his car (B) because he falls asleep at the wheel (A), but he would have crashed anyway because of black ice up ahead (A-2), there is an important sense in which the two crashes would not be identical, with respect to what the car hits, at what time, at what angle, etc. In fact, the two crashes would not be the same event at all. In this way, the argument of preemption seems to employ a sort of philosophical cheating. In a more rigorous and accurate way, preemption can be rephrased using the type/token distinction. Type of events B, such as "crashing the car," are indeed subject to preemption, but these are not a problem for Lewis. However, two token (specific) events can never be identical to each other, so that the same token of event can never have two causes such as in the above example. Preemption is thus a flawed argument, not only against Lewis's explanation of causation, but against all others as well.

There is, however, one criticism of the possible worlds approach that is, perhaps, a little more original and cannot be dismissed as easily as the two objections above. While possible worlds do indeed seem to be a method of truth evaluation for counterfactual conditionals, it is worth examining precisely how this is achieved. In the example above, "If Jimmy had driven, he would have crashed his car," we may check the validity of this statement using possible worlds. Assuming Jimmy did not drive (or crash his car) in our world, we take the nearest possible world – one in which Jimmy did indeed drive. Now, according to Lewis's method we check whether Jimmy indeed crashed his car in this world. But here lies the problem, for there does not seem to be any easy way to decide whether this is true. In fact, it appears once again that we must rely on factors such as Jimmy's driving record, road conditions, etc. Though things have become more complicated, we are back to using our intuitions regarding causation. Though intuition is a useful tool in philosophy, and one against which many theories and views can be tested, it is not at all helpful here – rather, determining a causal link using intuition is precisely what we sought to avoid. Lewis's entire proposal can be viewed as an attempt to remove ourselves from

---

---

intuition by finding a more formal, less subjective way to determine the presence of causation. Indeed, worse than making no progress, it seems that the possible worlds proposal actively hinders the effort. Starting from intuitions about causation, an entire complex proposal that is explanatorily and perhaps metaphysically demanding has been built, only to find that in the end it relies on precisely the same intuition to make the final judgment call. This critique appears devastating unless some response to it can be found.

Lewis's proposal, while flawed, is not doomed. Though it may seem from the critiques above that possible worlds overcomplicate the situation without moving past the use of intuition, it must be noted that causal dependence and causal chains are valuable insights, especially once the criticism from preemption is dismissed. In fact, the entire proposal could still be rescued with some suggestion of a way to move from our intuitions to a more mechanized explanation of causation once Lewis's paradigm has been employed and a nearest possible world found. Until such time, however, it appears that while it is an interesting suggestion, this account of causation does not advance our understanding of the phenomenon significantly. Causation proves once again to be a slippery philosophical opponent, and much remains to be investigated.

## NOTES

<sup>1</sup> See Hume, *Inquiries concerning Human Understanding and Concerning the Principles of Morals*.

<sup>2</sup> See Mackie, *The Cement of the Universe: A Study of Causation*.

<sup>3</sup> See Lewis, "Causation," 556-67; and Lewis, "Counterfactuals and Comparative Possibility," 418-46.

<sup>4</sup> Lewis, *Counterfactuals*, 1-3.

<sup>5</sup> Lewis, *On the Plurality of Worlds*, 1 - 92.

<sup>6</sup> Lewis, "Causation," 556-67.

**WORKS CITED**

- Hume, David. *Inquiries concerning Human Understanding and Concerning the Principles of Morals*, third edition, edited by L. A. Selby-Bigge. Oxford: Clarendon Press, 2006.
- Lewis, David. "Causation." *Journal of Philosophy* 70 (1973): 556-67.
- Lewis, David. *Counterfactuals*. Oxford: Blackwell Publishing, 1973.
- Lewis, David. "Counterfactuals and Comparative Possibility." *Journal of Philosophical Logic* 2 (1973): 418-46.
- Lewis, David. *On the Plurality of Worlds*. Oxford: Blackwell Publishing, 1986.
- Mackie, John. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press, 1974.