Philos Stud (2015) 172:2275–2278
DOI 10.1007/s11098-014-0409-6

Knowledge, counterfactuals, and determinism

Jeremy Goodman

Published online: 14 November 2014

Abstract Deterministic physical theories are not beyond the reach of scientific discovery. From this fact I show that David Lewis was mistaken to think that small counterfactual perturbations from deterministic worlds involve violations of those world's laws.

Keywords Knowledge · Counterfactuals · Determinism

Through scientific inquiry we can discover general truths about the physical world. Consider a scientist *S* who comes to know the truth of some physical theory *T* by inferring it as the best explanation of some of her observations. Since knowledge requires belief that is safe from error, we may conclude that *S* could not easily have come to falsely believe *T* by employing the very same reasoning that she actually used to come to know *T*. Now suppose that at some time long after her discovery a fair coin is tossed and lands tails. Of course, it could easily have landed heads instead. Had that happened, *S* would still have antecedently come to believe *T* by employing the very same reasoning that she actually used to come to know *T*. Since *S* could not easily have been in error, we may conclude that it is not the case that, had the coin landed heads instead, *T* would have been false.

We can formalize this argument as follows. Let *K* abbreviate '*S* knows *T*'; let *B* abbreviate '*S* comes to believe *T* by employing the very same reasoning that she actually used to come to know *T*' (where 'actually' picks out the world of the thought experiment); let *H* abbreviate 'The coin lands heads at the same time that it actually landed tails'; let $\ulcorner \mathcal{E}\varphi \urcorner$ abbreviate ⌜It could easily have been the case that φ⌝; let $\Box\!\!\rightarrow$ abbreviate the counterfactual conditional. Here is the argument:

J. Goodman (✉)
New York University, New York, USA
e-mail: goodman.jeremy@gmail.com

(1) K
(2) $\mathcal{E}H$
(3) $H \,\Box\!\!\rightarrow B$
(4) $K \supset \neg\mathcal{E}(B \wedge \neg T)$
(5) $(H \,\Box\!\!\rightarrow (B \wedge \neg T)) \supset (\mathcal{E}H \supset \mathcal{E}(B \wedge \neg T))$
(6) $((H \,\Box\!\!\rightarrow B) \wedge (H \,\Box\!\!\rightarrow \neg T)) \supset (H \,\Box\!\!\rightarrow (B \wedge \neg T))$
(7) Therefore, $\neg(H \,\Box\!\!\rightarrow \neg T)$

The argument is classically valid. (1), (2), and (3) follow from the description of the case. (4) is an instance of the principle that knowledge requires belief that is safe from error. (5) is an instance of the principle that what could easily have happened is closed under counterfactual implication. (6) is an instance of the principle that if two claims are counterfactually implied by a given claim, then so too is their conjunction. A defense of these principles is beyond the scope of this paper, but they are highly plausible.

Our conclusion (7) makes trouble for is Lewis (1979) influential account of counterfactuals. This is because his account entails a principle I will call the *Lewisian Lemma*—namely, that every true deterministic theory is such that, had the coin landed heads instead of tails, the theory would have been false. (A theory is *deterministic* just in case, for all times t and t', the theory together with a complete specification of the physical state of the universe at t entails a complete specification of the physical state of the universe at t'). Here is why.

Let t_0 be a time long before the coin is flipped, and let t_1 be the time when the coin lands tails. Let Φ_0 and Φ_1 be complete specifications of the physical states of the universe at t_0 and t_1 respectively. Let X be a true deterministic theory. Since X is deterministic, X must be inconsistent with the physical state of the universe at t_0 being Φ_0 and the physical state of the universe at t_1 not being Φ_1. Now suppose that the coin had landed heads at t_1. Had that happened, the physical state of the universe at t_1 would not have been Φ_1, since there could not be a difference in how the coin lands without there being a difference in the physical state of the universe. But according to Lewis, the physical state of the universe at t_0 would still have been Φ_0. His view therefore entails that, had the coin landed heads at t_1, two claims jointly inconsistent with X would both have been true. X would therefore have been false (since counterfactual implication is closed under entailment). Since X is an arbitrary true deterministic theory, we may generalize to establish the Lewisian Lemma.

Now suppose that the Lewisian Lemma were true. According to (7), it is not the case that T would have been false had the coin landed heads. So, by the Lewisian Lemma, T cannot be a true deterministic theory. Since S knows that T is true, T must be true. It follows that T is not deterministic. This conclusion generalizes, since the only thing we assumed about T was that S came to know it by inferring it as the best explanation of some of her observations. Moral: If Lewis's theory of counterfactuals is correct, then deterministic theories are thereby beyond the reach of scientific discovery.

But this consequence is absurd. Deterministic theories are obviously *not* thereby beyond the reach of scientific discovery. So we may legitimately stipulate that T is deterministic, since such cases are possible. Our case is then a counterexample to Lewis's theory of counterfactuals. It is not in general true that small counterfactual changes to the physical state of the universe at a time would make no difference to the physical state of the universe at much earlier times. In particular, it is not the case that the physical state of the early universe would have been just as it actually is had the coin landed differently from how it actually did.

This is not yet to establish that the physical state of the early universe *would* have been different from what it actually is had the coin landed differently from how it actually did. Establishing this further conclusion requires the premise:

(8) $H \mathbin{\Box\!\!\rightarrow} T$

according to which, had the coin landed heads, T would still have been true. Given (1)–(4), this premise follows from the following two claims:

(9) $((H \mathbin{\Box\!\!\rightarrow} B) \wedge (H \mathbin{\Box\!\!\rightarrow} \neg(B \wedge \neg T))) \supset (H \mathbin{\Box\!\!\rightarrow} T)$
(10) $(\mathcal{E}H \wedge \neg\mathcal{E}(B \wedge \neg T)) \supset (H \mathbin{\Box\!\!\rightarrow} \neg(B \wedge \neg T))$

(9) follows from the principle used to justify (6) above, together with the principle that tautologically equivalent formulas are inter-substitutable in the consequents of counterfactuals. (10) is an instance of the principle that anything that could easily have been the case counterfactually implies anything that could not easily have failed to be the case, which seems hard to deny.[1]

What makes T counterfactually robust? Presumably not the fact that S discovered it. Rather, for any natural way of fleshing out the case, T will be comprised of physical laws, and no claim with the status of physical law could easily have been false. (Since we have shown the Lewisian Lemma to be false, we are free to accept the pre-theoretically compelling claim that physical laws enjoy counterfactually robust truth whether or not they are deterministic.[2])

Suppose the physical laws *actually are* deterministic (as many philosophers of physics believe). It follows that, if the current physical state of the universe had been different from what it actually is in some way that it easily could have been, then the physical state of the distant past would have been different from what it actually was. Lewis hoped to avoid this startling conclusion. Surprisingly, his hopes are dashed by the epistemological banalities that knowledge requires belief that could not easily be mistaken and that deterministic theories can be known by scientific means.[3] Yet the assumption of physical determinism does *not* entail that, if

[1] (8) also follows directly from (7) given the (admittedly controversial) principle of conditional excluded middle, according to which anything that fails to counterfactually imply a given claim counterfactually implies that claim's negation.

[2] Dorr (ms) gives an argument for this conclusion closely related to my own, expanding on Dorr and Hawthorne (2014). He too argues that Lewis's theory of counterfactuals is false because it is possible to have a safe belief in a deterministic physical theory. Unlike me, he does not motivate this claim on epistemic grounds.

[3] An anonymous referee suggested that, granting that it is possible to *justifiably believe* a true deterministic theory, perhaps such beliefs never amount to knowledge. In reply: in addition to being an implausible concession to skepticism, this proposal has the unpalatable consequence that, if S had justification to believe it, she would have justification to accept the Moore-paradoxical conjunction 'T and I don't know T'.

the current physical state of the universe had been different from what it actually is in some way that it easily could have been, then the physical state of the distant past would have been *macroscopically* different from what it actually was. It therefore does not obviously threaten our practice of holding fixed macroscopic features of the past in ordinary counterfactual reasoning, or call into question our belief that in many respects the physical world could easily have been different from the way now it happens to be.[4]

Acknowledgments Thanks to Cian Dorr, Harvey Lederman, Bob Stalnaker, Tim Williamson and two Hanonymous referees for comments on previous versions of this paper.

References

Albert, D. Z. (2000). *Time and chance*. Cambridge, MA: Harvard University Press.

Dorr, C. (ms). Against counterfactual miracles.

Dorr, C., & Hawthorne, J. (2014). Semantic plasticity and speech reports. *Philosophical Review*, *123*(3), 281–338.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13*, 455–476.

Loewer, B. (2007). Counterfactuals and the second law. In H. Price & C. Richard (Eds.), *Causation, physics, and the constitution of reality* (pp. 293–326). Oxford: Oxford University Press.

Williamson, T. (2009). Probability and danger. *The Amherst Lecture in Philosophy*, *4*, 1–35.

4 Albert (2000), Loewer (2007), and Dorr (ms) argue on the basis of statistical mechanical considerations that our pre-theoretical judgments about ordinary counterfactuals are compatible with the counterfactually robust truth of deterministic physical laws. Williamson (2009) argues that, whether or not the laws of physics are deterministic, things could easily have been other than they actually are, noting that usually the best explanation of why a true belief fails to amount to knowledge is that the belief could easily have been false.