Comments on Katie Elliot's "Explanation Isn't Transitive"

DANIEL.HOEK@VT.EDU, CHAPEL HILL COLLOQUIUM, OCTOBER 2022

What a cool talk! I was particularly excited to see you bringing in the time travel example — circular explanations make my head spin a bit, as does time travel, but I happen to find that a pleasant sensation. Also I found this part of the talk illuminating because it shows that Katie's question is not just an abstraction for philosophers of science (and maybe grounding theorists). There's really something concrete at stake in the question whether explanation is transitive. For if we take the irreflexivity of explanation seriously (and I certainly do) then the transitivity of explanation would rule out the possibility of explanatory circles, and with it the possibility of time travel! So this question potentially has substantive implications for physics.

Just to get everybody's minds spinning in explanatory circles a little more, I want to start by expanding our arsenal a little bit and get you thinking about a different sort of example (less about *scientific* explanation per se, but it still seems relevant). Consider this dialogue:¹

A: "Why is it true that there are truths?"

- B: "Because there are truths."
- A: "Oh yeah?"
- B: "Yes, there are truths, because it is true that Katie just gave a talk."
- A: "And why is that true?"
- B: "Because she just gave a talk!"
- A: "I see I guess that checks out."

Now if you agree that this checks out, then it seems like we should be able to substitute any truth for *Katie just gave a talk* in this explanatory chain. So in particular, we could substitute the truth that *there are truths*. Uh, oh! Now we have a circle:

It is true that there are truths

↑

↓

There are truths

And so on pain of admitting that *there are truths* and *it is true that there are truths* both explain themselves, we need to ditch the transitivity of explanation! Grist to Katie's mill.

But now let me raise a point of disagreement. So far I have assumed both Katie and I are stalwart defenders of the irreflexivity of explanation (things do not explain themselves). But I actually think Katie is a bit of a false friend of irreflexivity. Or at any rate she isn't a true believer in the way that I am.

¹ This is a variant of a puzzle about grounding from Stephan Krämer's "A simpler puzzle of ground," *Thought* 2013. See also Peter Fritz, "On higher-order logical grounds," *Analysis* 2020.

The reason I say that will take a second to explain, and takes us back to Katie's main argument. Let me start with Katie's initial discussion of chancy explanations. She claimed that the meteorological conditions and laws explain why *there's a 70% chance of rain*, but not why *it will rain*. Moreover, she thinks *there's a 70% chance of rain* does explain why *it will rain*. Now I have a strong gut feeling that this last claim can't be right: the fact that *there's 70% chance of rain tomorrow* is not a genuine explanation for why *it will rain tomorrow*.

To see why I think that, consider this.² We know from Lewis that chances evolve over time, so that by the time that it's already raining, the chance of rain is now 100%. But evidently the fact that *there's a 100% chance it's raining now* does not explain the fact that *it's raining now*. That's because the purported explanans is insufficiently distinct from the explanandum. I think the same is true of *there's a 70% chance of rain* — it's still too close. A real explanation would have to cite distinct facts: *the wind is coming from this direction, there's an anticyclone over there yada yada* (I'm imagining these circumstances too take the chance up to 70%). Likewise, I think *There was 100% chance of Captain winning the Kentucky Derby* does not explain at all why *Captain won* — in fact this raises my puzzlement in this case! An explanation would be *There was a conspiracy and all the other horses had Nyquil dropped into their water troughs.*

For similar reasons I am inclined to reject Katie's raven example. In general, I think *necessarily p* is insufficiently distinct from *p* for it to be true that *necessarily p* explains *p*. Some examples:

- It is mathematically necessary that multiplication is commutative doesn't explain why multiplication is commutative.
- It is physically necessary for the pressure on the tank to rise as the temperature of the gas increases doesn't explain why the pressure on the tank rises as the temperature of the gas increases.
- It is metaphysically necessary that water is H₂O doesn't explain why water is H₂O (these identities probably can and need not be explained at all: a forteriori the fact that they are necessary does not explain them.)
- The laws of Newtonian mechanics explain many of phenomena, but they do not explain why the law of gravitation holds, or why *the change in motion of a body is proportional to the force impressed on it.* (Again, it may be that these laws don't need any explanation.)
- *It is a law that all ravens are black* does not explain why *all ravens are black*. (So even if Katie is right that the generalization explains that *ravens are hard to see at night* but the law does not, it's no counterexample to transitivity.)

Fundamentally, my take is that these are not explanations for the same reason that p does not

² The reason for my intuition is **not** that I think 0.7 is too low. I will accept explanations that fail to bring the chance up to 1 or anywhere close. E.g. I'd accept that *Captain is extremely fast, tenacious and well-trained* explains why *Captain won the race* (even though the explanans only brings Captain's odds up to 40%, say). And I'd go as far as accepting that *Jane bought 100 tickets* explains why *Jane won the million-ticket lottery* (although it only gives her a chance of 1 in 10 000). My issue is also not that chance facts cannot feature in explanations. E.g. chance facts explain distributions: *smokers have a greater chance of developing lung cancer* explains why *lung cancer is more prevalent in areas with more smokers*.

explain p: the explanans is insufficiently distinct from the explanandum. And that's why I said that Katie wasn't a true friend of irreflexivity. In the same way, I could never consider somebody a true friend of irreflexivity if they thought that *Jane and Jack are both in Paris* explains why *Jane is in Paris*: this may not be strictly of the form "p explains p", but the explanans is still insufficiently distinct from the explanandum.

Okay, so maybe Katie is not entirely opposed to self-explanation. What gives? In fact, if we're not all in on irreflexivity, why not just go whole hog over to the dark side and embrace self-explanations! Let's just run with that for a minute and see what happens: it'll be fun. Well, first of all we'd have to then give up the other strand of arguments against transitivity: the time travel-based non-transitive intuition, and also the other circular-explanation counterexample I was talking about before. All you have to do to restore transitivity here is to accept that the links in these circular explanatory chains explain themselves. And if we are to be friends of self-explanation, then we should definitely want to accept that. For you will surely not find more beautiful, intuitively compelling examples of self-explanation anywhere than in time travel stories and in these other explanatory circles.

Now if I'm honest, I must say this take on explanatory circles actually seems pretty attractive to me. In fact, I think that making an exception to irreflexivity in these cases may even be motivated by the mechanistic/causal conceptions of explanation that Katie nicely articulated. As I understand it, it is part of this conception that the explanation needs to reveal something about the structure of the larger mechanism in which the explanandum is embroiled. As Katie pointed out, this bars self-explanation in normal linear cases, because they do not show any of the wider mechanism. However, isn't it different in the time travel case? In the story, Katie's attempt to kidnap Tom Brady was an essential part of the mechanism which led to her attempt to kidnap Tom Brady. For it led to the victory of Tampa Bay, which caused Katie's enraged time travel, etcetera. So in this particular case, it seems to me that if it is properly understood, "Katie attempted to kidnap Tom Brady because she attempted to kidnap Tom Brady 'really does reveal something true and substantive about the (circular) structure of the wider mechanism in which the explanans is embroiled, and so it should count as a real explanation.

The more I've thought about it, the more I like this picture, both in the way it underpins my irreflexivity intuitions in linear cases, and in its analysis of why irreflexivity fails in the special case of explanatory circles. So tit looks like I don't have such inflexible irreflexivity intuitions after all...

Surprise surprise! It turns out that in the end, Katie and I were both false friends of irreflexivity — in fact, I am falser still because I actually end up denying it (assuming that there are explanatory circles, anyway). In any event, Katie and my basic intuitions about explanation are at bottom more similar than I thought they were at the start. Except for one trifling detail: I now think that explanation is probably transitive after all. But what gives?