TWO NEW COUNTEREXAMPLES TO THE TRUTH-TRACKING THEORY OF KNOWLEDGE

Tristan HAZE

ABSTRACT: I present two counterexamples to the recently back-in-favour truth-tracking account of knowledge: one involving a true belief resting on a counterfactually robust delusion, one involving a true belief acquired alongside a bunch of false beliefs. These counterexamples carry over to a recent modification of the theory due to Rachael Briggs and Daniel Nolan, and seem invulnerable to a recent defence of the theory against known counterexamples, by Fred Adams and Murray Clarke.

KEYWORDS: knowledge, truth-tracking, counterexamples

> In recent years Nozick's notion of knowledge as tracking truth has witnessed a revival. - Horacio Arló-Costa.[1]

Here I present two counterexamples to the truth-tracking theory of knowledge. As far as I have been able to tell, they are new. These counterexamples seem called-for in view of a recent defence and a recent modification of the theory (addressed below).

The simple version of Nozick's famous truth-tracking account runs as follows:[2]

S knows that p iff

1. p is true.

2. S believes that p.

3. If p weren't true, S wouldn't believe that p

4. If p were true, S would believe that p

---

[1] Horacio Arló-Costa, "Review of *Tracking Truth: Knowledge, Evidence and Science*, by Sherrilyn Roush," *Notre Dame Philosophical Reviews*, July 20, https://ndpr.nd.edu/news/25079-tracking-truth-knowledge-evidence-and-science/.

[2] Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981).

Tristan Haze

Counterexample 1: I have a deep-seated, counterfactually robust delusional belief that my neighbour is a divine oracle. He is actually a very reliable and truthful tax-lawyer. There is a point about tax law he has always wanted to tell me, *p*. One day, he tells me that *p*, and I believe him, because I believe he is a divine oracle. I would never believe him if I knew he was a lawyer, being very distrustful of lawyers.

In this case, it seems to me, I do not *know* that *p*: my belief rests on a delusion, albeit a counterfactually robust one. But it is true, I believe it, and my belief tracks the truth: if it were true, I would have believed it, and if it were false, I would not have believed it. (The lawyer, being reliable and truthful about tax law, would not have told me that *p* if *p* were not the case.)

Counterexample 2: My neighbour is a tax lawyer. Here, unlike in the previous counterexample, I have no delusional belief. It is my neighbour who is the strange one: for years, he has intently nurtured an eccentric plan to get me to believe the truth about whether *p*, where *p* is a true proposition of tax law, along with five *false* propositions about tax law. His intention to do this is very counterfactually robust. He moves in next door and slowly wins my trust. One day, he begins to regale me with points of tax law. He asserts six propositions: *p* and five false ones. I believe them all.

It seems to me that I do not know that *p* in this case either. But I believe it, it is true, and my belief tracks the truth: if *p* were the case, I would have believed it, and if *p* were not the case, I would not have believed it (remember, the tax lawyer has long been anxious that I believe the truth about whether *p*).

These counterexamples can easily be seen to carry over to Nozick's more complicated method-relativized version of the account, since there is only one method in question in each case. That version goes *via* an account of knowing-by-a-method which runs as follows:[3]

S knows, via method (or way of knowing) M, that *p* iff

1. *p* is true

2. S believes, via method M, that *p*

3. If p weren't true, and S were to use M to arrive at a belief whether (or not) *p*, then S wouldn't believe, via M, that *p*

4. If p were true, and S were to use M to arrive at a belief whether (or not) *p*, S would believe, via M, that *p*.

---

[3] Nozick, *Philosophical Explanations*, 179.

They also carry over straightforwardly to the recent account of Rachael Briggs and Daniel Nolan,[4] which replaces counterfactuals with dispositions. (Their account was designed to deal with cases where the truth-tracking account undergenerates. Here, it *over*generates.)

Furthermore, they are unaffected by a recent defence of the truth-tracking account, due to Fred Adams and Murray Clarke,[5] against already-known putative counterexamples; these ones seem importantly different, and nothing Adams and Clarke say carries over to them, at least in any way I have been able to discern.

Note also that there is no objection to these counterexamples to be had in protesting that beliefs based on delusions cannot be knowledge, or that unreliable methods cannot lead to knowledge – to insist on such things for putative cases of knowledge is simply to depart from the type of account under discussion.

The two counterexamples are quite different from each other. I put both forward because each seems interesting in its own way, and because two counterexamples to a false theory are better than one. (I find both convincing, but perhaps some readers will accept one and not the other.)

If I were more of an optimist I would conclude by saying that perhaps now we can finally relieve this tired old theory from being a contender, and instead learn from it a useful negative lesson about knowledge.[6]

---

[4] Rachael Briggs and Daniel Nolan, "Mad, Bad and Dangerous to Know," *Analysis* 72, 2 (2012): 314–316.

[5] Fred Adams and Murray Clarke, "Resurrecting the Tracking Theories," *Australasian Journal of Philosophy* 83, 2 (2005): 207–221.

[6] Thanks to John Turri, Fred Adams and Murray Clarke for helpful correspondence.