Why is knowledge faster than (true) belief?

BBS Commentary on "Knowledge before belief," by Jonathan Phillips, Wesley Buckwalter, Fiery Cushman, Ori Friedman, Alia Martin, John Turri, Laurie Santos, and Joshua Knobe

Evan Westra

Abstract: Phillips and colleagues convincingly argue that knowledge attribution is a faster, more automatic form of mindreading than belief attribution. However, they do not explain what it is about knowledge attribution that lends it this cognitive advantage. I suggest an explanation of the knowledge-attribution advantage that would also help to distinguish it from belief-based and minimalist alternatives.

One of the key claims of the target article is that reasoning about states of knowledge is faster and more automatic than reasoning about states of belief. While Phillips and colleagues provide a range of evidence to support this claim, they do not explain what it is about knowledge attribution that makes it more efficient than belief attribution. Filling in these details will be crucial to explaining how knowledge attribution actually works and would also help to distinguish the proposed framework from nearby alternatives.

One way for the authors to explain the knowledge-attribution advantage would be to adopt a *minimalist* approach, and suggest that knowledge-based mindreading deploys representations of non-propositional relations that hold between agents and states of affairs – something analogous to Burge's notion of *sensing* (Burge, 2018) or Butterfill and Apperly's notion of a *registration* (Butterfill & Apperly, 2013). However, as proponents of these minimalist models have been careful to point out, this kind of mindreading does not actually enable agents to reason about propositional attitudes; rather, they enable agents to *track* mental states like belief without representing them as such. Since knowledge is also a propositional attitude, this means that minimal mindreading could not support genuine knowledge attribution. At most, it would enable an agent to extensionally track factive states without representing them *as knowledge*. If the account described in the target article aims for more than this, then a minimalist approach will not do.

A better approach to explaining the knowledge-attribution advantage would be to start by looking at the processing demands stemming *false*-belief attribution, the paradigmatic example of propositional attitude reasoning. Famously, false-belief attribution requires mindreaders to generate and maintain two mutually inconsistent, decoupled representations of the world, which places inherent demands executive functions like working memory and inhibitory control (Fizke et al., 2014; Schuwerk et al., 2014). If knowledge attribution involved a similar decoupling process, albeit one where the attributed representation is consistent with the mindreader's own primary, first-personal representation, then this might explain the knowledge-attribution advantage: while it involves the attribution of full-blown propositional attitudes, the contents of knowledge attributions do not conflict with the way the mindreader sees the world, which places fewer demands on their executive resources. However, this picture sounds perilously close to how one might describe *true*-belief attribution. If the knowledge-attribution advantage were due solely to this consistency in attributed contents, then it would seem that the entire model could be easily redescribed in non-factive, doxastic terms without any real loss in explanatory power.

Careful to distinguish their model from such non-factive alternatives, Phillips and colleagues point to evidence showing that mindreading in Gettier-like cases where an agent has true beliefs but not knowledge is actually quite difficult, both for children and for nonhuman primates (Fabricius et al., 2010; Horschler et al., 2019). These data indicate that reasoning about mere true belief is surprisingly inefficient, despite the fact that it does not require agents to generate inconsistent representations of the world. Thus, whatever it is that explains the knowledge-attribution advantage, it is not its similarity to mere true-belief attribution. But while this response provides support for the knowledge-attribution framework, it does not bring us any closer to answering our central question. Instead, we are left with a new puzzle: what makes knowledge attribution different from true-belief attribution?

Here, we must consider the different cognitive demands that arise when reasoning about the mind of a knower and the mind of a mere true believer. While both forms of mindreading involve attributing a mental state that matches reality, the way that the mindreader must represent the link between that mental state and the world is not the same in the two cases. The representations that we attribute to the knower are securely bound to the contents of our own primary representations of the world. This is most obvious in cases of shared knowledge – for example, when two people both have unobstructed perceptual access to the same event. From such a shared epistemic position, maintaining a model of another agent's mind is easy, because updates to our own primary representations of the world seamlessly carry over to our knowledge attributions. This is much different than the kind of decoupling involved in false-belief reasoning, where we must actively quarantine off our representations of the other agent's mental states so that they can be updated separately. In knowledge attribution, in contrast, representations of the knower's mental states are tightly *coupled* with our own primary representations (Westra & Nagel, under review).

In the case of the mere true believer, the link between mind and the world is much less secure. Although the contents we attribute in this case *happen* to align with the way we take the world to be, the weakness of their epistemic position means that the possibility of misalignment lurks nearby. In this sense, the mere true believer is not so different from the false believer: both kinds of case demand a kind of *epistemic vigilance* (Sperber et al., 2010), an alertness to the potential for error that is unnecessary in ordinary cases of knowledge-attribution (Nagel, 2019). It is thus unsurprising that reasoning about mere true beliefs appears to be so cognitively demanding, since it requires a level of decoupling between mind and world similar to what we see in false-belief attribution.

One prediction that follows from this analysis is that whenever knowledge attributions *do* require heightened levels of epistemic vigilance, they will not be much more efficient than the corresponding belief attributions. One context where this might occur would be in Frege cases, where we must treat an agent as having knowledge of a referent under one mode of presentation but not under another – for example, when we represent Lois Lane as knowing that Superman flies, but not as knowing that Clark Kent flies (c.f. Rakoczy et al., 2015). In these cases, the processing demands (and hence, the speed) of belief-attribution and knowledge attribution should be equally demanding.

References:

Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review*, *125*(3), 409–434. https://doi.org/10.1037/rev0000091

Butterfill, S., & Apperly, I. (2013). How to Construct a Minimal Theory of Mind. *Mind and Language*, *28*(5), 606–637.

Fabricius, W. V., Boyer, T. W., Weimer, A. A., & Carroll, K. (2010). True or false: Do 5-year-olds understand belief? *Developmental Psychology*, *46*(6), 1402–1416. https://doi.org/10.1037/a0017648

Fizke, E., Barthel, D., Peters, T., & Rakoczy, H. (2014). Executive function plays a role in coordinating different perspectives, particularly when one's own perspective is involved. *Cognition*, *130*(3), 315–334. https://doi.org/10.1016/j.cognition.2013.11.017

Flavell, J. H., Everett, B. a., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103. https://doi.org/http://dx.doi.org/10.1037/0012-1649.17.1.99

Horschler, D. J., Santos, L. R., & MacLean, E. L. (2019). Do non-human primates really represent others' ignorance? A test of the awareness relations hypothesis. *Cognition*, *190*(April), 72–80. https://doi.org/10.1016/j.cognition.2019.04.012

Nagel, J. (2019). Epistemic Territory. *Proceedings and Addresses of the American Philosophical Association*, *93*, 67–86.

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit Theory of Mind Is Even More Unified Than Previously Assumed: Belief Ascription and Understanding Aspectuality Emerge Together in Development. *Child Development*, *86*(2), 486–502. https://doi.org/10.1111/cdev.12311

Schuwerk, T., Schecklmann, M., Langguth, B., Döhnel, K., Sodian, B., & Sommer, M. (2014). Inhibiting the posterior medial prefrontal cortex by rTMS decreases the discrepancy between self and other in Theory of Mind reasoning. *Behavioural Brain Research*, *274*, 312–318. https://doi.org/10.1016/j.bbr.2014.08.031

Sperber, D., Fabrice, C., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0017.2010.01394.x

Westra, E., & Nagel, J. (under review). *Mindreading in conversation*.