Belief vs. knowledge: An epic battle, but no clear victor

Daniel Lassiter
Stanford University

Abstract The knowledge-first approach is attractive and consistent with a wide variety of evidence. So is the opposing belief-first picture. I explain why the target article's criticisms of the latter fail, and argue that the outcome is a stalemate.

The target article (Phillips, Buckwalter, Cushman, Friedman, Martin, Turri, Santos & Knobe 2020) argues that the study of Theory of Mind in cognitive science should treat knowledge, rather than belief, as the 'basic' epistemic concept — conceptually more basic, evolutionarily prior, and appearing earlier in human development. While the article makes a compelling case that the knowledge-first approach is compatible with a wide variety of existing evidence, its efforts to undermine the competing belief-first approach are less convincing. Most or all of the evidence offered is equally compatible with both approaches, and the conflict that the authors set up seems to be an empirical stalemate. Let me explain.

A simplified application of the belief-first picture to adult human theory of mind is structured around two axes of distinction: opinionated and unopinionated states, and — among opinionated states — true and false states. This yields a 3-way classification: adult humans represent other agents as having, variously, no belief, a true belief, or a false belief on a certain topic. As the target article reviews, there is evidence that certain populations — great apes, children below the age of 4, and autistic patients — have difficulty in tasks where success would require representing another agent as having and acting upon a false belief. In the case of human children, Perner (1991) influentially attributes this difficulty to a lack of metarepresentational abilities that would needed to distinguish the child's own model of the world from another's mistaken model. As a result, children under 4 generally mistake false-belief states for either true-belief or no-belief states. In contrast, children between 2 and 4 have little difficulty distinguishing opinionated from unopinionated states of belief. Knowledge as a separate category develops later, as children become sensitive to sources of evidence (Perner 1991: ch.7). Modulo the eventual development of knowledge as a separate category, parallel phenomena arise with great apes and autistic patients, though there is debate as to whether the same theoretical interpretation is appropriate.

The knowledge-first picture is strikingly similar in structure. The parallel 3-way classification is now between ignorance, knowledge, and failed knowledge (i.e., false belief; Williamson 2000). The target article shows that the phenomena just outlined can be redescribed, perhaps more elegantly, if we suppose that great apes, small children, and autistic patients are able to make the knowledge-ignorance distinction but are unable to maintain separate representations of beliefs that do not count as knowledge. On this reading, failure to correctly predict actions based on false beliefs is due to treating the false-belief state as one of either knowledge or ignorance.

The problem is that these two accounts are very difficult to distinguish empirically. The target article attempts to do so by noting that there is much evidence that all three relevant populations

'represent knowledge', but little evidence that they 'represent belief'. However, in each of the three cases the key evidence shows something different and more equivocal. For instance, when arguing that chimpanzees do not 'represent belief', the authors adduce evidence that chimpanzees do not represent *false belief*. But this is no refutation of the belief-first picture: it is built in from the start, as described above. The same problematic style of argumentation occurs in the discussion of evidence around small children and autistic patients. In each case, the interpretation offered is quite reasonable: the evidence indicates a distinction between knowledge and ignorance with no corresponding distinction between knowledge and mere belief. However, it is equally compatible with the belief-first picture, where it would indicate distinction between no-belief and true-belief states, with no separate category for false-belief states.

Because of their structural parallelism, the choice between knowledge- and belief-first pictures cannot be made on the basis of which distinctions are being made at a coarse level. Perhaps, though, it could be made by asking fine-grained questions about the character of the true-and-opinionated category: true belief in the belief-first picture, and knowledge in the knowledge-first picture. Evidence from Gettier cases could in principle make it possible to choose, since they involve true beliefs that do not constitute knowledge. The target article cites a handful of studies with small children and great apes involving Gettier-like scenarios and construes them as evidence for the knowledge-first picture. However, this interpretation is somewhat tendentious. For instance, as Horschler, Santos & MacLean (2019) describe, the results involving change of location ('Sally-Ann') tasks with great apes can be explained more parsimoniously by supposing that apes are merely tracking whether another ape had perceptual access to the most recent event involving the item in question.

Similarly, the failure of children aged 4-6 to predict others' behavior on the basis of accidentally true beliefs may be better interpreted as an instance of a more general phenomenon: just as they begin to succeed on false-belief tasks, they start to fail even extremely simple, non-Gettiered true-belief tasks. In a series of experiments, Oktay-Gür & Rakoczy (2017); Rakoczy & Oktay-Gür (2020) show that this surprising failure does not arise in a non-verbal task, and that it can be modulated by simplifying the pragmatics of the task in various ways. Other authors have attributed this behavior to a perceptual access heuristic similar to that described above for great apes (though this would not explain sensitivity to pragmatic manipulations). In either case, children's and apes' apparent sensitivity to Gettier-like scenarios may turn out to be attributable to independent factors that are readily intelligible within the belief-first picture.

None of this casts doubt on the correctness of the knowledge-first picture, which is theoretically elegant and compatible with a wide range of empirical evidence. But the belief-first picture is also compatible with the available evidence, and the outcome of the skirmish is thus much less lopsided than the target article suggests. We can, however, hope that the authors' clear exposition of the knowledge-first position will inspire empirical studies that may eventually allow us to discern which position is correct.

References

Horschler, Daniel J, Laurie R Santos & Evan L MacLean. 2019. Do non-human primates really represent others' ignorance? a test of the awareness relations hypothesis. *Cognition* 190. 72–80.

Oktay-Gür, Nese & Hannes Rakoczy. 2017. Children's difficulty with true belief tasks: Competence deficit or performance problem? *Cognition* 166. 28–41.

Perner, Josef. 1991. *Understanding the representational mind*. The MIT Press.

Phillips, Jonathan, Wesley Buckwalter, Fiery Cushman, Ori Friedman, Alia Martin, John Turri, Laurie Santos & Joshua Knobe. 2020. Knowledge before belief. *Behavioral and Brain Sciences* 1–37.

Rakoczy, Hannes & Nese Oktay-Gür. 2020. Why do young children look so smart and older children look so dumb on true belief control tasks? an investigation of pragmatic performance factors. *Journal of Cognition and Development* 21(2). 213–239.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.