Behavioral and Brain Sciences

cambridge.org/bbs

Knowledge before belief

Jonathan Phillips[a], Wesley Buckwalter[b], Fiery Cushman[c], Ori Friedman[d], Alia Martin[e], John Turri[f], Laurie Santos[g], and Joshua Knobe[h]

[a]Program in Cognitive Science, Department of Psychological and Brain Sciences and Department of Philosophy, Dartmouth College, Hanover, NH 03755, USA; [b]Department of Philosophy, Institute for Philosophy and Public Policy, George Mason University, Fairfax, VA 22030, USA; [c]Department of Psychology, Harvard University, Cambridge, MA 02138, USA; [d]Department of Psychology, University of Waterloo, Waterloo, ON, N2L 3G1, Canada; [e]School of Psychology, Victoria University of Wellington, Wellington, 6012, New Zealand; [f]Philosophy Department and Cognitive Science Program, University of Waterloo, Waterloo, ON N2L 3G1 Canada; [g]Department of Psychology, Yale University, New Haven, CT 06520, USA and [h]Program in Cognitive Science, Department of Philosophy, Yale University, New Haven, CT 06520, USA.

jonathan.s.phillips@dartmouth.edu; http://phillab.host.dartmouth.edu/
wesleybuckwalter@gmail.com; https://wesleybuckwalter.org/
cushman@fas.harvard.edu; http://cushmanlab.fas.harvard.edu/
friedman@uwaterloo.ca; https://sites.google.com/view/uwaterloocclab
alia.martin@vuw.ac.nz; https://vuwbabylab.com/
john.turri@gmail.com; https://john.turri.org/
laurie.santos@yale.edu; https://caplab.yale.edu/
joshua.knobe@yale.edu; https://campuspress.yale.edu/joshuaknobe/

Target Article

Cite this article: Phillips J, Buckwalter W, Cushman F, Friedman O, Martin A, Turri J, Santos L, Knobe J. (2021) Knowledge before belief. *Behavioral and Brain Sciences* **44**, e140: 1–75. doi:10.1017/S0140525X20000618

Target Article Accepted: 03 September 2020
Target Article Manuscript Online: 08 September 2020
Commentaries Accepted: 30 December 2021

Key words:
belief; factivity; false belief; knowledge; knowledge-first; theory of mind

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 17) and an Author's Response (p. 67). See bbsonline.org for more information.

Abstract

Research on the capacity to understand others' minds has tended to focus on representations of *beliefs*, which are widely taken to be among the most central and basic theory of mind representations. Representations of *knowledge*, by contrast, have received comparatively little attention and have often been understood as depending on prior representations of belief. After all, how could one represent someone as knowing something if one does not even represent them as believing it? Drawing on a wide range of methods across cognitive science, we ask whether belief or knowledge is the more basic kind of representation. The evidence indicates that nonhuman primates attribute knowledge but not belief, that knowledge representations arise earlier in human development than belief representations, that the capacity to represent knowledge may remain intact in patient populations even when belief representation is disrupted, that knowledge (but not belief) attributions are likely automatic, and that explicit knowledge attributions are made more quickly than equivalent belief attributions. Critically, the theory of mind representations uncovered by these various methods exhibits a set of signature features clearly indicative of knowledge: they are not modality-specific, they are factive, they are not just true belief, and they allow for representations of egocentric ignorance. We argue that these signature features elucidate the primary function of knowledge representation: facilitating learning from others about the external world. This suggests a new way of understanding theory of mind – one that is focused on understanding others' minds in relation to the actual world, rather than independent from it.

1. Introduction

Research on how people understand each other's minds tends to focus in particular on how people attribute *beliefs* (e.g., Baron-Cohen, 1997; Call & Tomasello, 2008; Dennett, 1989; Nichols & Stich, 2003; Onishi & Baillargeon, 2005; Saxe & Kanwisher, 2003). However, people also have other ways of understanding each other's minds, including attributing *knowledge*. That is, instead of asking "What does this person believe?," we can ask "What does this person know?" Knowledge attribution has received far less attention than has been devoted to belief. A simple Google Scholar search, for example, demonstrates approximately an order of magnitude more papers that focus on tests for representations of belief than representations of knowledge in theory of mind.[1]

The reasons for this focus on belief are both methodological and historical. Because beliefs can be false, they provide a convenient method for testing whether an agent's mind is being represented independently from one's own representation of the external world. Moreover, historically, beliefs have been taken to be among the most conceptually basic mental states. People use many different concepts to make sense of the way other people understand the world, including the concepts of guessing, assuming, suspecting, presupposing, fully anticipating, and so forth. But the concept of belief may be more fundamental than any of these, and people's use of all of these other concepts may depend on their ability to represent beliefs. Knowledge may well be in the same camp as these other mental states. That is, people's ability to represent knowledge may ultimately depend on a more fundamental capacity to represent

© The Author(s), 2020. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

Jonathan Phillips is an Assistant Professor in Cognitive Science at Dartmouth College. His research employs the methods of psychology, philosophy, linguistics, and computer science to better understand how people represent and reason about non-actual possibilities, and the role that these representations play in higher-order cognition, including theory of mind, causal reasoning, linguistic communication, and moral judgment.

Wesley Buckwalter is an Assistant Professor in the Department of Philosophy and a Faculty Fellow in the Institute for Philosophy and Public Policy at George Mason University. His research explores several topics at the intersection of philosophy and cognitive science pertaining to social cognition and communication involving knowledge, belief, responsibility, bias, and moral judgment.

Fiery Cushman is the John L. Loeb Associate Professor of Social Sciences in the Department of Psychology at Harvard University. He studies human moral judgment and decision-making and its constituent parts, including value-guided learning and decision-making, mental-state reasoning, and causal reasoning.

Ori Friedman is a Professor of Psychology at the University of Waterloo. His research explores conceptual understanding and development in children and adults and seeks to understand how people quickly and flexibly deploy conceptual knowledge to understand the world. His recent research investigates topics including how children and adults reason about ownership of property; how children and adults judge whether events are possible, and how they decide which beliefs count as knowledge.

Alia Martin is a Senior Lecturer in Psychology at the Victoria University of Wellington and director of the VUW Infant and Child Cognition Lab. She received her PhD in Psychology from Yale University in 2014. Her research focuses on the developmental and phylogenetic origins of mental-state representation, and how human infants and children use mental-state inferences in the context of communication and other social behaviors.

John Turri is Canada Research Chair in Philosophy and Cognitive Science at the University of Waterloo. He studies concepts, judgments, and practices central to commonsense cognition and communication.

Laurie Santos is Professor of Psychology and Cognitive Science and Head of Silliman College at Yale University. Her research explores the evolutionary origins of the human mind by studying the cognitive capacities of nonhuman animals, specifically nonhuman primates and domesticated dogs. She is the winner of the Stanton Prize from the Society for Philosophy and Psychology for outstanding contributions to interdisciplinary research and an American Psychological Association Distinguished Scientific Award for an Early Career Contribution to Psychology. She has been recently voted one of Popular Science Magazine's Brilliant 10 Young Minds and one of Time Magazine's "Featured Campus Celebrities."

Joshua Knobe is a professor at Yale University, appointed both in the Program in Cognitive Science and in the Department of Philosophy. Most of his research is in the field of experimental philosophy. Thus, much of his research uses the sorts of experimental methods familiar from cognitive science to address the sorts of conceptual questions familiar from philosophy. In his research so far, he has explored questions about the ordinary concepts of intentional action, causation, essence, the true self, free will, and, of course, knowledge.

belief. This way of understanding the relationship between knowledge and belief has been widespread in the philosophical literature (for a review, see Ichikawa & Steup, 2017) and may further justify the focus on belief in research on theory of mind.

Surprisingly, empirical research offers little support for this way of understanding the relationship between knowledge and belief. Instead, most of the empirical evidence to date points in the opposite direction: the capacity to attribute knowledge is more basic than the capacity to attribute belief. We review the evidence for this conclusion from a wide range of fields using a diversity of methodological approaches: comparative psychology, developmental psychology, social psychology, clinical psychology, and experimental philosophy. This evidence indicates that nonhuman primates attribute knowledge but not belief (Section 4.1), that the capacity to attribute knowledge arises earlier in human development than the capacity to attribute belief does (Section 4.2), that implicit knowledge attributions are made more automatically than belief attributions (Section 4.3), that the capacity to represent knowledge may remain intact in patient populations, even when belief representation is disrupted (Section 4.4), and that explicit knowledge attributions do not depend on belief attributions (Sections 5.1 and 5.2) and are made more quickly than belief attributions (Section 5.3). Together these converging lines of evidence indicate that knowledge, rather than belief, is the more basic mental state used to represent other minds.

This abundance of evidence naturally gives rise to a further question of why representations of knowledge play such a basic role in the way we understand others' minds. We argue that the set of features that are specific to knowledge representations suggest a promising answer: a primary function of knowledge representations is to guide learning from others about the external world. This presents a new view of how to think about theory of mind – one that is focused on understanding others' minds in relation to the actual world, rather than independent from it.

2. How should we understand *knowledge*?

Given our aim, an important initial question concerns what we mean by *knowledge*. At the broadest level, our proposal will be to start by treating knowledge as the ordinary thing meant when people talk about what others do or do not "know." While there is some disagreement about how to best make sense of the ordinary meaning (Ichikawa & Steup, 2017), there is also near-universal agreement on a number of essential features that knowledge has, and we take those as the signature markers of the kind of representation we are interested in.

Specifically, we focus on four features that are essential to knowledge:

2.1. Knowledge is factive

You can believe whatever you like, but you can only know things that are true. Also, when you represent others' knowledge, you can only represent them as knowing things that you take to be true. If you do not think it is true that the moon landing was faked, you cannot rightly say, "Bob knows the moon landing was faked." You would instead have to say, "Bob believes the moon landing was faked." Representations of knowledge *can only* be employed when you take the content of the mental state to be true (Kiparsky & Kiparsky, 1970; Williamson, 2000).

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

2.2. Knowledge is not just true belief

The capacity for attributing knowledge to others is not the same as a capacity for attributing true belief. Many cases of true belief are not knowledge. In the most widely discussed kinds of cases, someone's belief ends up being true by coincidence. For example, suppose that John believes that a key to his house is in his pocket. Unfortunately, that key fell out of his jacket as soon as he put it in. However, another house key, which he forgot about, is in the other pocket of the jacket he is wearing. In such a case, John has a belief that there is a house key in his pocket, and that belief happens to be true. Intuitively though, it seems wrong to describe John as *knowing* that there is a house key in his pocket (Gettier, 1963; Machery et al., 2017; Starmans & Friedman, 2012).

2.3. Others can know things you do not

While you can't represent others as knowing things that are false, you can represent others as knowing things you do not know. You can say, for example, "Suzy knows where to buy an Italian newspaper" even if you do not know where to buy an Italian newspaper (Karttunen, 1977; Phillips & George, 2018). Accordingly, the capacity to represent knowledge involves the capacity to represent two different kinds of ignorance. On the one hand, you can represent yourself as knowing more than others do ("altercentric ignorance"), but at the same time, you can also represent others as knowing more than you do ("egocentric ignorance") (Nagel, 2017; Phillips & Norby, 2019).

2.4. Knowledge is not modality-specific

While knowledge may be gained through different forms of perception (seeing, hearing, feeling, and so on), representations of knowledge are not tied to any particular sense modality; knowledge is more general than that. Moreover, knowledge can be gained without perception, for example, by logical inference. So attributing knowledge goes beyond merely representing what others see, hear, feel, and so on.

These four essential features of knowledge helpfully distinguish it from other kinds of mental-state representation. For example, representations of belief lack the feature of being factive, whereas representations of seeing or hearing, unlike knowledge, are modality-specific. Hence, our plan throughout will be to focus on instances of mental-state representations that have these four signature features of knowledge. We will then ask how such a capacity for knowledge representation relates to the ability to represent others' beliefs.

3. Two views about knowledge and belief

We will begin by setting out two broad views about how to understand the relationship between belief attribution and knowledge attribution. We start by considering these views at a purely theoretical level, without looking at empirical data or at prior investigations of these questions. The remainder of the paper then turns to existing empirical and theoretical studies to determine which of these two views is better supported by the evidence.

When considering the role of knowledge and belief in theory of mind, it will be important to distinguish between two closely related questions. The first asks whether one of these ways of representing others' minds is more basic than the others (e.g., whether one preceded the other in evolutionary history, or is present earlier in human development, or is computed more quickly). The second asks more specifically whether the less basic attribution depends in some way on the more basic one. The answers to these two questions will obviously be connected in an important way, since it would be difficult to see how the more basic representation could depend on some less basic one. At the same time, though, the less basic representation might not depend in any way on the more basic one; they could be entirely independent of each other. With these two questions in mind, let us consider two ways of understanding the role of belief and knowledge in theory of mind.

3.1. View 1: Belief attribution is more basic

A view familiar to philosophers holds both that belief is the more basic mental state, and that representations of others' knowledge *depend* in a critical way on representations of their beliefs. Applying this picture to psychology, people may make knowledge attributions by (a) making a simpler belief attribution and (b) also going through certain further cognitive processes that give rise to a representation that goes beyond merely attributing a belief.

If we are wondering about John's mental states regarding where his keys are, we would default to representing John's mind in terms of where John *believes* his keys are (for specific proposals on how we might do this, see, e.g., Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017; Goldman, 2006; Gopnik & Wellman, 1992; Gordon, 1986; Leslie, Friedman, & German, 2004). Then, to determine whether or not John *knows* that the keys are in his pocket, we would additionally need to determine whether his belief also meets the additional criteria required for knowledge.

For a more specific proposal about these criteria, one might turn to the rich philosophical literature, which has explored the ways in which the concept of knowledge goes beyond the concept of belief (Armstrong, 1973; Dretske, 1981; Goldman, 1967; Lewis, 1996; Nozick, 1981; Sosa, 1999). In what follows, we will not be concerned with those details. Our concern is rather with the broad idea that belief is the more basic way in which we represent others' minds and that representations of knowledge depend on representations of belief. If this idea is right, then we would expect there to be a set of processes that produce comparatively simple representations of what others believe (see, e.g., Rose, 2015). This view aligns well with the proposal that the capacity for belief attribution may in fact be a part of core cognition or otherwise innate, and present extremely early in human development (see, e.g., Kovács, Téglás, & Endress, 2010; Leslie et al., 2004; Onishi & Baillargeon, 2005; Stich, 2013).

On this view, belief is more basic than knowledge in both senses. It is more basic in the sense that we would expect, for example, representations of belief to be computed more quickly than representations of knowledge. Further, belief is more basic than knowledge in that knowledge representations may depend on representations of belief.

3.2. View 2: Knowledge attribution is more basic

An alternative view is that *knowledge* is the more basic way in which we make sense of others' minds. Rather than representing what others know by first representing what they believe, people may have a separate set of processes that give rise to some comparatively simple representation of what others know. Such a representation clearly would not involve calculations of belief.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

The signature features of knowledge illustrate ways in which knowledge representations are substantially more limited than belief representations. For example, knowledge is factive, so if you do not think that it is true that John's keys are in his pocket, then you certainly cannot represent him as knowing that his keys are in his pocket. Moreover, even if you do think it is true that John's keys are in his pocket, but John's belief about the location of his keys only happens to be right as a matter of coincidence, then once again you cannot represent him as knowing that his keys are in his pocket.

In contrast, belief representations are more flexible. You can easily represent John as believing that his keys are in his pocket, or in any other location: his car, his shoe, or Timbuktu. In fact, representations of belief do not seem to be restricted in any way: in principle, you could represent John as believing completely arbitrary propositions. Hence, the set of things you could represent someone as knowing is necessarily smaller than the set of things you could represent them as believing (Phillips & Norby, 2019). These differences between knowledge and belief representation suggest that there may be some relatively simple representation of what others know and some comparatively complex representation of others' beliefs.

While the knowledge-as-more-basic view denies that knowledge representations depend on representations of belief, it does not make any commitment as to whether the converse holds. Even if knowledge representations are simpler than belief representations, the capacity to represent someone as believing something need not depend on the capacity to represent them as knowing something. These two ways of understanding others' minds may simply be independent.

3.3. Shifting focus

The origin of the focus on beliefs in theory of mind research is easily traceable to Premack and Woodruff's paper, "Does the chimpanzee have a theory of mind?" in *Behavioral and Brain Sciences* (1978). In the commentaries to that target article, a number of philosophers argued that for both theoretical and empirical reasons, better evidence for a capacity for theory of mind would show that chimpanzees represented beliefs, and in particular, false beliefs (Bennett, 1978; Dennett, 1978; Harman, 1978).

The idea that belief may be among the most basic mental representations is common within philosophy, and can be clearly observed in epistemological discussion of whether knowledge may be understood as some, augmented, version of justified true belief (e.g., Armstrong, 1973; Chisholm, 1977; Dretske, 1981). On this picture, belief is understood as the basic concept and knowledge is taken to be comparatively complex, that is, belief that has various other properties (being true, justified, and so on). However, philosophers have long questioned whether belief should actually be understood as more basic than knowledge, and recently, have become increasingly excited about the possibility that knowledge is the more basic notion (Nagel, 2013; 2017; Williamson, 2000; see Carter, Gordon, & Jarvis, 2017 for a collection of recent approaches; see Locke, 1689/1975; Plato, 380BCE/1997, for earlier related discussions). On this approach, knowledge should not be analyzed in terms of belief or other concepts but should be understood as basic and unanalyzable.

Just as many philosophers have rethought the commitment to belief being more basic than knowledge, we think it is time for cognitive scientists to revisit their continued emphasis on belief as to the more central or basic theory of mind representation. In fact, we think that the empirical research from cognitive science provides overwhelming evidence in favor of thinking that it is knowledge attribution, rather than belief attribution, that is, the more basic theory of mind representation.

In the sections that follow, we ask what cognitive science research reveals about this issue. We first review how the tools of cognitive science allow one to test claims about which types of representations are basic. We then examine what each of these empirical methods illustrates about the basicness of belief and knowledge representations. The first group of methods we examine are primarily nonlinguistic and do not involve specifically asking subjects what they think about what someone "knows" or "believes." Rather, these methods involve operationalizing knowledge and belief within an experimental protocol and then investigating whether subjects' behavior demonstrates sensitivity to either kind of representation. The second group of methods actually employ the terms "knowledge" and "belief" and ask what people's use of these words can tell us about the basicness of the corresponding concepts. As detailed in the following sections, one finds a strikingly similar story across all of these highly varied methods.

4. Is belief understanding more basic than knowledge understanding?

A central task in cognitive science is to uncover and describe the most basic functions of human cognition – what "core" parts of the mind make up the foundation that we use to develop more complex, experience-dependent ways of thinking about the world. To answer this question, scientists have marshaled a set of empirical tools that give insight into these more basic aspects of human cognition.

First, researchers have tested which aspects of human cognition are more basic by examining the *evolutionary history* of different kinds of capacities, testing which aspects of human understanding are shared with closely related primate species and thus are likely to be evolutionarily foundational. Second, researchers have investigated which capacities emerge earliest in *human ontogeny*, and thus do not require much experience. Third, researchers have examined which cognitive capacities in adult humans operate *automatically*, and thus occur independently of conscious initiation. (The underlying logic here is not that if a capacity is automatic, then it is basic since there are capacities that are automatic but not basic, such as reading. Rather, the idea is that if a capacity were a very basic one, it would be surprising if it did not operate automatically.) Finally, researchers have examined which capacities are more basic by testing *special populations* that are unique either in terms of their experiences or in terms of neural variation (e.g., brain lesions, autism spectrum disorder [ASD], and so on). Capacities that are more basic tend to be conserved across populations despite radical differences in experiences or deficits in other cognitive processes. Taken together, this set of tools can provide important interdisciplinary insights into the question of which aspects of the mind are the most basic and form the foundation for other more complicated abilities.

To see these empirical tools of research, consider another domain in which scholars have also wondered which sorts of capacities are the more basic: the domain of human numerical understanding. Our ability to think about complex numerical concepts requires a host of sophisticated capacities, many of

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

which require the kinds of complex computational abilities that only adult humans possess. But which parts of this numerical understanding are basic and require some additional complex processing? Using the combined empirical approaches described above, researchers have determined that at least two aspects of our adult human numerical understanding – the capacity to represent large sets of objects approximately and the capacity to represent small numbers of individual objects exactly – appear to be basic (see reviews in Carey, 2009; Feigenson, Dehaene, & Spelke, 2004; Spelke, 2004). First, closely related primates appear to have both the capacity to exactly represent the number of objects in a small set (e.g., Santos, Barnes, & Mahajan, 2005) and the capacity to make approximate guesses about large numbers of objects (e.g., Jordan & Brannon, 2006). Second, human infants also appear to begin life with both of these capacities: they can track small numbers of objects (e.g., Wynn, 1992) and make quick approximate guesses about large numbers of objects (e.g., Xu & Spelke, 2000). Third, adult humans appear to perform both of these sorts of tasks automatically: we automatically subitize small numbers of objects exactly (e.g., Trick & Pylyshyn, 1994) and seem to automatically guess which of two large sets has approximately more objects (e.g., Barth, Kanwisher, & Spelke, 2003). Finally, researchers have identified special populations in which participants' understanding of approximate numbers are preserved despite radical differences in experience (e.g., the Munduruku, an indigenous group of people who live in the Amazon River basin, Dehaene, Izard, Spelke, & Pica, 2008). These combined developmental, comparative, automaticity, and special population findings make a convincing case that the ability to enumerate objects approximately and track small numbers of objects exactly is more basic than other aspects of human number cognition (see review in Carey, 2009). For more discussion of the analogy between number cognition and theory of mind, see Apperly and Butterfill (2009), Spelke and Kinzler (2007), and Phillips and Norby (2019).

With this analogy in mind, let us return to the first question we posed about the relationship between knowledge and belief: Is belief attribution or knowledge attribution more basic in the sense described above?

In the following sections, we review what these same four empirical tools – comparative cognition research, developmental studies with infants and young children, studies of automatic processing in adult humans, and deficits in special populations – reveal about the foundational aspects of our understanding of other minds. What emerges from examining research with each of these empirical tools is a picture that is just as consistent as the one observed for number representation. All four of these different empirical tools suggest that representations of what others know are more basic than representations of what others believe.

4.1 Knowledge and belief in nonhuman primates

The first empirical tool that we marshal is comparative studies examining what nonhuman primates understand about others' minds. Much research over the past few decades has investigated how a number of primate species think about the minds of others and this study has given us a relatively clear picture of what different primates understand about others' mental states (see reviews in Call & Santos, 2012; Drayton & Santos, 2016). Considering this now large body of research, we can ask whether we see an understanding of others' *beliefs* emerging as a more phylogenetically foundational aspect of mental-state reasoning in primates.

First off, do any nonhuman primates actually represent others' beliefs? Looking to our closest primate relatives, the great apes, one finds mixed evidence for an understanding of beliefs. Three recent sets of studies support the conclusion that chimpanzees and some other great apes can represent others' false beliefs (Buttelmann, Buttelmann, Carpenter, Call, & Tomasello, 2017; Kano, Krupenye, Hirata, Tomonaga, & Call, 2019; Krupenye, Kano, Hirata, Call, & Tomasello, 2016). However, there is also some reason for caution when interpreting these results. Researchers have continued to debate whether these findings are better explained by lower-level processing (Heyes, 2017; Kano, Krupenye, Hirata, Call, & Tomasello, 2017; Kano et al., 2019; Krupenye, Kano, Hirata, Call, & Tomasello, 2017), simpler representations that do not involve false belief (Buttelmann et al., 2017; Tomasello, 2018), and even whether the anticipatory-looking paradigms used in this research have reliably demonstrated theory of mind in humans (Dörrenberg, Rakoczy, & Liszkowski, 2018; Kulke, Wübker, & Rakoczy, 2018; Schuwerk, Priewasser, Sodian, & Perner, 2018). In brief, these studies provide some initial evidence that great apes can represent false beliefs, but additional research continues to be warranted (Martin, 2019).

At the same time, many other published studies suggest that apes fail to represent others' beliefs across a range of tasks (Call & Tomasello, 1999; Kaminski, Call, & Tomasello, 2008; Krachun, Carpenter, Call, & Tomasello, 2009; O'Connell & Dunbar, 2003). In one study, for example, Kaminski et al. (2008) explored whether chimpanzees could understand that a competitor had a false belief about the location of a hidden food item. Kaminski et al. used a design in which subject chimpanzees competed with a conspecific for access to contested foods (for the first of such tasks, see Hare, Call, Agnetta, & Tomasello, 2000). Chimpanzees did not distinguish between a condition where the competitor had a false belief and a condition where the competitor was ignorant (the food was taken out and then replaced in the same container in the competitor's absence). They were more likely to go for high-quality food in both of these conditions than in a knowledge condition where the competitor had seen where the food ended up. These results suggest that chimpanzees fail to account for false beliefs in competitive tasks, but they have no trouble distinguishing knowledge from ignorance (for similar results, see Call & Tomasello, 2008; Krachun et al., 2009).

In comparison to the mixed evidence one finds for representations of belief in great apes, the picture is clear when it comes to great apes' representations of knowledge: great apes have shown robust success in representing what others know and acting in accordance with those representations (Bräuer, Call, & Tomasello, 2007; Hare et al., 2000; Hare, Call, & Tomasello, 2001; Hare, Call, & Tomasello, 2006; Kaminski et al., 2008; Karg, Schmelz, Call, & Tomasello, 2015; Krachun et al., 2009; Melis, Call, & Tomasello, 2006; Whiten, 2013). Collectively, these studies suggest that great apes can track what others know in competitive tasks, even though they often fail to track others' beliefs in those same tasks (see, e.g., Call & Tomasello, 2008; MacLean & Hare, 2012).

Importantly, research on nonhuman primates has also investigated more distantly related primates, like monkeys, which provides insight into which capacities may have evolved even longer ago. The evidence regarding monkeys is even more unequivocal. To date, there is no evidence that monkeys

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

understand other individuals' beliefs, even when tested on tasks that human infants have passed (Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011; Martin & Santos, 2016). Research on mental-state understanding in human infants often uses looking-time measures (e.g., Kovács et al., 2010; Onishi & Baillargeon, 2005). When these same techniques are applied to monkeys, they do not show evidence of representing false beliefs (Martin & Santos, 2014) or using them to predict behavior (Marticorena et al., 2011). In one study, monkeys watched an event in which a person saw an object moved into one of two boxes and then looked away as the object moved from the first box into the second box. Once the person had a false belief about the location of the object, monkeys appeared to make no prediction about where she would look; they looked equally long when the person reached either of the two locations (Marticorena et al., 2011; see also Martin & Santos, 2014 for similar results on a different task). These findings suggest that primates more distantly related to humans than great apes fail to represent beliefs, indicating that the human ability to represent beliefs may actually be phylogenetically recent.

In spite of monkeys' difficulty in tracking others' beliefs, there is a large body of work demonstrating that monkeys can understand what others know (Drayton & Santos, 2016; Martin & Santos, 2016). Rhesus monkeys, for example, understand that they can steal food from a person who cannot see the food (Flombaum & Santos, 2005) or who cannot hear their approach toward the food (Santos, Nissen, & Ferrugia, 2006). Moreover, when monkeys' understanding of others' knowledge states are tested using looking-time measures, researchers again observe a dissociation in monkeys' understanding of knowledge and belief. For example, when rhesus monkeys see a person watching an object going into one of two locations, they look longer when that person reaches the incorrect location than the correct location, suggesting that they expect people to search correctly when they know where an object is (Marticorena et al., 2011). These findings suggest that more phylogenetically distant monkey species succeed in tracking others' knowledge states even though they fail to understand others' beliefs.

4.1.1. Do nonhuman primates actually represent knowledge?

An essential further question is whether the research on nonhuman primate theory of mind actually provides evidence regarding knowledge representations specifically, rather than something else, such as a representation of perceptual access. To answer this further question, we need to ask whether there is evidence that the theory of mind representations observed in nonhuman primates carry the signature features that are unique to knowledge. A number of studies provide evidence that this is the case (see Nagel, 2017, for a complementary perspective).

First, there is evidence that nonhuman primates can represent egocentric ignorance; that is, they can represent someone else as knowing something they do not know. For example, in a competitive task involving obtaining food from one of two containers, chimps and bonobos were placed in a position such that they could see that their human competitor could see which container the food was placed in, but they could not see where the food was placed (Krachun et al., 2009). The positions of the two containers were then switched in clear sight of both the subject and their human competitor. When subjects searched for food, they demonstrated a marked preference for taking food from the container their competitor reached for, suggesting that they represented the competitor as knowing where the food was even if they did not.[2] Critically, in a minimally different false-belief condition where the containers switched positions whereas the competitor was not watching, chimps and bonobos (unlike 5-year-old children) were unable to recognize that they should search in the container the competitor was not reaching for (Krachun et al., 2009).

Second, there is evidence that apes and monkeys fail to represent others' *true beliefs* in cases where they have no trouble representing others' knowledge (Horschler, Santos, & MacLean, 2019; Kaminski et al., 2008). Specifically, these studies included conditions where food was placed in one of the two opaque containers in clear sight of both the experimenter and the nonhuman primate. Then, after the experimenter's line of sight was occluded, the food was removed from the container but then put directly back in the same container where it was originally placed. Under such conditions, the experimenter should have a true belief about the location of the food (since it did not actually change locations), but not knowledge (Gettier, 1963). Strikingly, under these conditions, nonhuman primates failed to predict that the experimenter would act on the basis of the true belief. In contrast, they have little trouble making the correct predictions in matched conditions where the experimenter could be represented as having knowledge because they saw the removal and replacement of the food (Horschler et al., 2019; Kaminski et al., 2008).

Finally, there is evidence that the knowledge representations found in nonhuman primates are not modality-specific. For example, both chimpanzees and rhesus macaques make the same inferences about others' knowledge based on auditory and visual information (Melis et al., 2006; Santos et al., 2006). Moreover, recent research studies also suggest that both chimpanzees and macaques can attribute inferential knowledge to others that cannot be solely based on perceptual access (Drayton & Santos, 2018; Schmelz, Call, & Tomasello, 2011).

Taken as a whole, the lesson from comparative research is that the capacity to represent others' beliefs may be evolutionarily newer than the capacity to represent others' knowledge. There is mixed evidence that great apes track others' beliefs. Nonetheless, there is clear evidence that great apes are able to track others' knowledge. Going yet a further step across the evolutionary tree, there is clear evidence that monkeys can track others' knowledge in a variety of contexts but not others' beliefs. In short, primate research studies to date suggest that the capacity to think about others' knowledge predates an ability to represent others' beliefs.

4.2 Knowledge and belief in human development

4.2.1. Knowledge and belief in infancy

Just as studies of nonhuman primates can provide evidence about which cognitive capacities are evolutionarily more foundational, so too can studies of preverbal infants demonstrate which cognitive capacities are *developmentally prior*. In the last two decades, a growing body of work using non-verbal methods provides evidence that preverbal infants have the capacity to represent the mental states of others.

The current evidence of non-verbal belief representation in early infancy is, at this point, unequivocally mixed. A number of studies have suggested that infants reason about an agent's actions in terms of her *beliefs* by15 months of age or earlier (Buttelmann, Carpenter, & Tomasello, 2009; Kovács et al., 2010; Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007; Träuble, Marinović, & Pauen, 2010). At the same time, there

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

have been a number of compelling plausible proposals for how key behavioral patterns can be explained without a genuine ability for belief representation (Burge, 2018; Butterfill & Apperly, 2013; Heyes, 2014a; Priewasser, Rafetseder, Gargitter, & Perner, 2018). Further, other researchers have argued that some of these looking-time patterns may actually reflect representations of knowledge rather than belief (Wellman, 2014).

More obviously concerning though, recent attempts at replicating or extending the key pieces of empirical data have been unsuccessful (e.g., Dörrenberg et al., 2018; Grosse Wiesmann et al., 2018; Kammermeier & Paulus, 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018). At this point, the field is largely in disagreement about whether there is good evidence for a capacity for belief representation in human infants. Rather than taking a side in this debate, however, we simply want to point out that whichever way this debate turns out, the ability to represent knowledge seems to replicably precede an ability to represent beliefs.

There is uncontroversial evidence that infants can appreciate how others' knowledge shapes their actions from at least six months of age. First, six-month-old infants are sensitive to the role of an agent's current or prior perceptual knowledge in constraining that agent's actions toward objects. For example, six-month-old infants usually assume that an agent who reaches for object A over B prefers object A and will continue to reach for that object in the future. However, if the agent's view of object B is occluded during the initial reaching demonstration, infants do not infer this preference; indeed, they make no prediction about the agent's future behavior when the agent has not seen both options. In this way, infants recognize that an agent's knowledge of her surroundings affects agent's future behavior (Luo & Johnson, 2009). Six-month-olds also make similar inferences based on what an agent has seen previously. For example, after observing an interaction between a "communicator" who prefers object A over B and a naive "addressee," they do not expect the addressee to provide the communicator's preferred object. However, infants at this age do expect the addressee to provide the communicator's preferred object when the addressee was present and watching during the communicator's initial preference display, or when the communicator uses an informative vocalization during the interaction (a speech sound). Hence, six-month-olds seem to recognize some of the conditions under which an individual will become knowledgeable about information (Vouloumanos, Martin, & Onishi, 2014). These two examples and many others (e.g., Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; Luo, 2011; Luo & Baillargeon, 2007; Meristo & Surian, 2013) suggest that within the first year of life infants reason about agents' actions in terms of what those agents know and do not know and how others' knowledge states shape their actions. Yet there is comparatively little evidence that infants before the second year of life have an ability to represent others' beliefs (see Kovács et al., 2010).

An important aspect of these studies (as well as the nonhuman primate research) is that they required researchers to rely on tasks with solely *nonlinguistic responses*. To complement this study, we next turn to consider studies that directly ask young children about what others "know" or "believe" and consider the developmental trajectory of knowledge and belief in these tasks.

4.2.2. Knowledge and belief in young children

Research on preverbal infants' understanding of others' epistemic states provides evidence about which cognitive capacities emerge earliest in life and may serve as the foundation for other later-emerging capacities demonstrated in verbal reports. While some uncertainty remains about the relationship between these two sets of capacities (see, e.g., Apperly, 2010; Baillargeon, Scott, & He, 2010; Carruthers, 2013, 2016), research suggests that the developmental sequence observed in preverbal infants bears a striking similarity to the sequence of development found by researchers studying verbal reports in preschool-aged children. Once again, the capacity for identifying and employing representations of knowledge precede those of belief.

One simple way to track the emergence of the concepts of knowledge and belief in childhood is to consider children's naturally occurring language production and comprehension. Studies of children's early language use suggest that children typically grasp factive mental-state terms first, and more specifically understand the mental-state verb "know" before "think." Toddlers, for example, use "know" in their own utterances before they use "think" (e.g., Bartsch & Wellman, 1995; Bloom, Rispoli, Gartner, & Hafitz, 1989; Shatz, Wellman, & Silber, 1983; Tardif & Wellman, 2000). While there remains some debate about how children understand these terms (Dudley, 2018), there is good evidence that preschoolers grasp the relative certainty conveyed by "know" before they grasp this for "think" (Moore, Bryant, & Furrow, 1989). Moreover, children make systematic errors when using nonfactive mental-state terms like "think," which suggest that they may first interpret these terms as factive (e.g., misinterpreting "think" as "know") (de Villiers, 1995; de Villiers & de Villiers, 2000; de Villiers & Pyers, 1997; Johnson & Maratsos, 1977; Lewis, Hacquard, & Lidz, 2012; Sowalsky, Hacquard, & Roeper, 2009; though see Dudley, Orita, Hacquard, & Lidz, 2015). To illustrate, one error that young children often make is to deny belief ascriptions whenever the agent's belief does not meet the standards of knowledge, for example, the belief is false. That is, when children are asked whether an agent "thinks" something, their patterns of answers indicate that they are actually answering a question about whether or not an agent "knows" something. Finally, corpus analyses of toddlers' uses of the term "know" also reveal an early-emerging understanding of knowledge in that they use these terms to both signal their own ignorance and request that knowledgeable others fill in gaps in their understanding (Harris, Bartz, & Rowe, 2017a; Harris, Ronfard, & Bartz, 2017b; Harris, Yang, & Cui, 2017c).

Another large body of research has experimentally varied the information an agent has acquired and then asked children to make inferences about what the agent "knows" or "thinks." For instance, in a typical task assessing inferences of knowledge, an agent is either shown the contents of a closed container, or is not shown, and children are then asked whether the agent knows what is in the container. Here, success requires attributing knowledge when the agent saw the contents and attributing ignorance when the agent did not. Similarly, in a typical false-belief task, an agent sees that an object is in one location but does not see it get moved to another location, and children are asked where the agent thinks the object is. To succeed here, children must indicate that the agent thinks the object is in the first location, even though children themselves know it is in the second location.

Findings from studies using these verbal measures suggest that children succeed in inferring knowledge states before they successfully infer belief states. Whereas successful attribution of knowledge states often emerges when children are aged 3 (e.g., Pillow, 1989; Pratt & Bryant, 1990; Woolley & Wellman, 1993),

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

successful attribution of false belief typically occurs only when they are 4 or older (Grosse Wiesmann, Friederici, Disla, Steinbeis, & Singer, 2017; see Wellman, Cross, & Watson, 2001 for a meta-analysis of findings from false-belief tasks). Particularly compelling evidence for this pattern comes from studies that have used a battery of theory of mind tasks developed by Wellman and Liu (2004). These studies show that most children succeed in inferring knowledge states before they succeed in attributing false belief (Mar, Tackett, & Moore, 2010; Tahiroglu et al., 2014) and that this developmental pattern is stable across a variety of populations, including deaf children and children with autism (Peterson, Wellman, & Liu, 2005), and children from non-Western cultures (Shahaeian, Nielsen, Peterson, & Slaughter, 2014; Shahaeian, Peterson, Slaughter, & Wellman, 2011; Wellman, Fang, Liu, Zhu, & Liu, 2006).

Considering young children's capacity for making explicit, verbal judgments of knowledge and belief, one sees a familiar pattern emerge. Much like in non-verbal tasks, young children succeed in verbal tasks that require facility with representations of knowledge before they succeed in tasks that require facility with representations of belief.

4.2.3. Is this really the development of knowledge representations?

Again, a critical question is whether the research we have reviewed on theory of mind in human development actually provides evidence that infants and young children are representing knowledge rather than something else, such as perceptual access or simply true belief. That is, do we see the signature features of a genuine capacity for representing *knowledge*?

One important piece of evidence comes from studies asking whether infants have a capacity for egocentric ignorance: are they able to represent that others know something that they do not? Behne, Liszkowski, Carpenter, and Tomasello (2012) had an experimenter hide an object in one of the two boxes in a way that ensured that infants could not infer which box the object was in. When the experimenter then pointed to one of the two boxes, infants searched for the object in the location pointed to, suggesting that they understood that the experimenter knew something they did not (Behne et al., 2012). Along similar lines, Kovács, Tauzin, Téglás, Gergely, and Csibra (2014) provided evidence that 12-month-old infants' pointing is used to query others who they take to be more knowledgeable than they are (Kovács et al., 2014). Specifically, they found that infants exhibited a tendency to point in cases where the experimenter was likely to provide knowledge that the infant did not have (compared to a case where the experimenter was likely to share information that the infant already knew). Moreover, Begus and Southgate (2012) demonstrated that 16-month-old infant's interrogative pointing is sensitive to the previously demonstrated competence of potential informers, suggesting that this pointing demonstrates a genuine desire to learn what others *know*, rather than merely believe (see Stenberg, 2013, for convergent evidence with 12-month-old infants). Collectively, this study provides evidence that infants have an early-emerging capacity to represent others as knowing something that they do not know – a signature property of knowledge representation.

Continuing later into development, this capacity is also evident in 3-year-olds' explicit attributions of knowledge (e.g., Birch & Bloom, 2003; Pillow, 1989; Pratt & Bryant, 1990; Woolley & Wellman, 1993) and their decisions about who to ask for help (Sodian, Thoermer, & Dietrich, 2006). Indeed, children's ability to represent others as knowing more than themselves can be seen as underwriting the important and well-studied development of young children's trust in testimony (see Harris, Koenig, Corriveau, & Jaswal, 2018, for a recent review). From the perspective of a human infant seeking to learn from others, being able to represent others as knowing something you do not know is critical for understanding who to learn from.

Second, studies show that young children fail to correctly attribute true beliefs if they fall short of knowledge (Fabricius, Boyer, Weimer, & Carroll, 2010; Fabricius & Imbens-Bailey, 2000; Oktay-Gür & Rakoczy, 2017; Perner, Huemer, & Leahy, 2015). These studies have employed scenarios that are similar to "Gettier" cases within epistemology. To illustrate, children in one study were told about a boy named Maxi who knows that his mother placed his chocolate in the red cupboard, but then while Maxi is gone, his sister takes the chocolate out of the cupboard and after eating some, considers putting it in the green cupboard. However, in the end, she decides to just put it back in the red cupboard (Fabricius et al., 2010). In this situation, Maxi has a justified true belief about his chocolate being in the red cupboard, but he is not properly described as *knowing* that his chocolate is in the red cupboard (Gettier, 1963). The striking finding is that even at an age where they can clearly represent knowledge (4- to 6-year-olds), children fail to correctly predict where Maxi will look for the chocolate when his true belief falls short of genuine knowledge. In contrast, when the paradigm is minimally changed such that it no longer involves a "Gettier" case but can be solved with genuine knowledge representations, young children no longer have any difficulty correctly predicting where the agent will look (Oktay-Gür & Rakoczy, 2017). In short, children fail to correctly predict others' behavior when their true beliefs fall short of knowledge.

Third, there is good evidence that mental-state representation in infants is not completely explained by modality-specific perceptual access relations such as seeing-that or hearing-that. Infants make the same inferences about what others know based on both auditory and visual information, suggesting that there is some common, modality-independent representation of what others know (Martin, Onishi, & Vouloumanos, 2012; Moll, Carpenter, & Tomasello, 2014). Additionally, infants attribute knowledge based on nonperceptual inferences that the agent should make. For example, Träuble et al. (2010) showed 15-month-old infants an agent who is either facing the display (and thus has perceptual access to a ball changing locations) or is not facing the display but manually adjusts a ramp causing the ball to change locations (and thus can make a physics-based inference about the ball's changed location). In both cases, infants regarded the agent as having knowledge of the changed location of the ball and distinguished these cases from ones where the agent did not have reason to infer that the ball changed locations (Träuble et al., 2010). In fact, there is striking evidence that young children (3- to 5-year-olds) are actually surprisingly bad at tracking the modality through which agents gain knowledge of an object (O'Neill, Astington, & Flavell, 1992; Papafragou, Li, Choi, & Han, 2007). Young children will, for example, assume that agents who have gained knowledge of an object through only one sense modality (e.g., touch) also have gained knowledge typically acquired through other sense modalities (e.g., what the object's color is). This kind of error suggests that children are relying primarily on a general capacity for representing others as simply knowing (or not knowing) things about the world rather than modality-specific perceptual access, such as seeing-that or feeling-that.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

In sum, we find remarkably good evidence that the early-emerging theory of mind capacity has the signature features of genuine *knowledge* representation.

4.3 Automatic theory of mind in human adults

A third empirical way to test which cognitive capacities are more basic is to ask which processes operate *automatically* in human adults – that is, which capacities operate even when you do not want them to and continue to function even when the representation being computed is completely irrelevant (or even counterproductive) to the task at hand. To return to the previous example of number cognition, consider the difference between seeing 27 dots appear on a screen and seeing three dots appear on the same screen. When 27 dots appear, whether or not you represent the *exact* number of dots on the screen just depends on whether you intentionally decide to engage in the controlled process of counting the number of dots. You could spontaneously decide to count the number of dots, but you could just as easily decide not to. The capacity giving rise to representations of 27 dots is not automatic. Representations of three dots work differently. The processes involved in representations of three dots operate *automatically*: you could *not* realize that there are three dots, even if you wanted to (Dehaene & Cohen, 1994; Kaufman, Lord, Reese, & Volkmann, 1949; Trick & Pylyshyn, 1994). A growing body of literature has investigated the question of whether representations of knowledge and belief are *automatic*.

4.3.1 Current evidence for automatic belief representation

First, consider the evidence for whether people automatically compute belief representations. The most common approach has been to ask participants to make a judgment in response only to the information that they have seen, whereas at the same time, systematically varying the information that was presented to another (irrelevant) agent in the experiment (Apperly & Butterfill, 2009; Kovács et al., 2010; Low & Watts, 2013; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Surtees, Apperly, & Samson, 2016a; Surtees, Samson, & Apperly, 2016b). Researchers could then ask whether participants were automatically (i.e., mandatorily) calculating the mental states of the irrelevant agent by asking whether participants' responses were influenced by the information available to this other agent.

The evidence uncovered by research using this kind of paradigm has provided a relatively clear answer: there is little evidence that human adults automatically represent others' beliefs and some positive evidence that they do not. One prominent study has largely served as the primary piece of evidence for automatic belief representation (Kovács et al., 2010). Importantly, however, further research demonstrated that the paradigm used in these studies suffered from subtle confounds in the timing of a critical attention check, and once these confounds were controlled for, or simply removed, the results no longer suggested that participants automatically calculated others' beliefs (Phillips et al., 2015). Apart from this prominent piece of evidence, there are also a few other studies that have argued in support of automatic belief representation (Bardi, Desmet, & Brass, 2018; El Kaddouri, Bardi, De Bremaeker, Brass, & Wiersema, 2019; van der Wel, Sebanz, & Knoblich, 2014), and considerable evidence that strongly suggests that belief representation is *not* automatic (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Kulke et al., 2019; Low & Edwards, 2018; Surtees, Butterfill, & Apperly, 2012; Surtees et al., 2016a, 2016b).

Complementary evidence comes from studies asking whether representations of others' beliefs are computed when attentional resources or executive functions are taxed. This body of work provides clear evidence that representing what others believe requires deliberative attention and executive function (Apperly, Back, Samson, & France, 2008; Apperly, Samson, & Humphreys, 2009; Dungan & Saxe, 2012; Qureshi, Apperly, & Samson, 2010; Schneider, Lam, Bayliss, & Dux, 2012). To illustrate with one example, Dungan and Saxe (2012) had participants view videos in which an agent either formed knowledge of the location of an object or instead formed a false belief about the location of the object. They then asked participants to predict where the agent would look for the object when they were under various forms of cognitive load, using both verbal and non-verbal shadowing. When participants needed to represent the agent as having a false *belief* about the object's location, they made systematic errors in their predictions of where the agent would look. No such errors were observed when they could simply represent the agent as knowing the object's location.

In summary, the current state of the evidence suggests that not only are belief representations *not* automatic, but they rely on domain-general executive resources. Representations of beliefs work much more like representations of 27 dots than representations of three dots.

4.3.2. Current evidence for automatic knowledge representation

In contrast, one finds intriguing evidence that human adults may automatically represent what others know. In one study, Samson et al. (2010) showed that participants took into account what others knew, even in cases where it was counterproductive to the task they were completing. Participants viewed a room with various numbers of dots on two opposing walls, and their task was to indicate the number of dots they saw on the walls. Critically, however, there was also an avatar standing in the middle of the room, facing only one of the walls, such that on some trials, the participant saw more dots than the avatar did. On trials where the number of dots seen by the avatar and participant conflicted, participants tended to make errors in a way that suggested they were automatically encoding the number of dots the avatar saw, and that this representation was conflicting with their representation of the number of dots on the walls (despite the fact that the avatar was completely irrelevant to the task they were currently performing on these trials). This research, along with a number of subsequent studies (Surtees & Apperly, 2012; Surtees et al., 2016a, 2016b) collectively suggest that when we automatically encode others' mental states, we represent the things that they know through clear perceptual access (sometimes referred to as "Level-1" perspective taking, see Flavell, 1978, 1992). At the same time, however, there is continuous debate about whether these findings reflect the genuine theory of mind representations or simply lower-level processing required by the task (Heyes, 2014b), with some researchers providing evidence for attentional confounds (Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014), and others providing empirical evidence against the proposed confounds (Furlanetto, Becchio, Samson, & Apperly, 2016; Gardner, Bileviciute, & Edmonds, 2018; Marshall, Gollwitzer, & Santos, 2018).

Rather than attempting to adjudicate this debate, we instead want to step back and consider what this approach to investigating knowledge and belief has uncovered. There are two possibilities. One is that existing evidence from these paradigms shows

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

that humans are capable of automatically attributing knowledge but are not capable of automatically attributing beliefs. The other is that, despite the initial evidence, the experimental paradigms that have been employed so far are not well-suited to demonstrate the existence of an underlying capacity for automatic theory of mind in general, and new paradigms need to be developed.

4.3.3. New horizons for the automatic theory of mind

The research reviewed in the previous sections on nonhuman primates and human cognitive development provides reason to expect that the capacity for knowledge representation is more cognitively basic than belief. Moreover, many basic cognitive capacities – those shared with close nonhuman primate relatives and early emerging in human development – also tend to operate automatically in humans. Hence, there is some reason to expect that adult humans may indeed have the capacity to automatically represent others' knowledge. If this is right, we should further expect such an automatically functioning capacity for knowledge representation to exhibit the same set of signature features of knowledge representations. Specifically, we would expect this capacity to (i) only support *factive* representations, (ii) *not* support representations of true belief when they fall short of knowledge, (iii) allow you to represent someone else is knowing something you do not, and (iv) not be tied to any particular sense modality. To the best of our knowledge, none of these four features have been directly examined in this study on the automatic theory of mind, and thus point to exciting new avenues for future study as work on this topic presses forward.

4.4. Evidence from patient populations

The other tool that cognitive scientists use to determine which capacities are more basic is to ask which capacities are preserved in people who suffer from various cognitive impairments. The underlying rationale is that the more basic capacities tend to be preserved in patient populations. While there has not yet been a great deal of research that has specifically investigated which theory of mind capacities may be preserved across different patient populations, it is worth considering what the existing evidence may reveal about the capacities for representing knowledge and belief.

The most well-studied patient population in the theory of mind research are people with ASD. Research looking at the theory of mind in people with ASD has found that they often have difficulties in correctly representing others' beliefs (Baron-Cohen, 1997; Baron-Cohen, Leslie, & Frith, 1985; Frith, 2001; Moran et al., 2011; Schneider, Slaughter, Bayliss, & Dux, 2013; Senju, Southgate, White, & Frith, 2009). While there is also some evidence that young children with ASD have some difficulty with inferences about knowledge, representations of knowledge fare better than representations of belief when directly compared (Baron-Cohen & Goodhart, 1994; Leslie & Frith, 1988; Perner, Frith, Leslie, & Leekam, 1989; Pratt & Bryant, 1990). For example, in studies that tracked participants' eye movements during true and false-belief tasks, researchers found that the eye movements of people with ASD differ from controls when the agent has a false belief, they actually do not differ when the agent simply has knowledge (Schneider et al., 2013).

Similarly, a number of studies have investigated how people with ASD differ from typically developing people in terms of the ability for "Level 1" and "Level 2" theory of mind. In general, studies using Level 1 tasks (involving calculations of whether or not someone has perceptual or epistemic access to something) have found that people with ASD often perform just, as well as typically developing controls (Baron-Cohen, 1989; Hobson, 1984; Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Reed & Peterson, 1990; Tan & Harris, 1991). In contrast, studies involving Level-2 tasks (involving taking someone's perspective even though it contradicts one's own) have shown that people with ASD often perform much less than typically developing controls (Hamilton, Brindley, & Frith, 2009; Leslie & Frith, 1988; Reed & Peterson, 1990; Yirmiya, Sigman, & Zacks, 1994).

Taken together, this research suggests that the capacity for representing others' beliefs is disrupted in patient populations, whereas the capacity to represent what other people see or know remains comparatively preserved. This difference in disruptions of knowledge and belief provides evidence that the capacity for knowledge representation is more basic than belief representation. Not only is knowledge more basic in that it is simpler, but knowledge representation clearly does not depend on belief representation, since representations of knowledge are preserved despite disruptions of belief representations.

4.5. Summary

The tools that cognitive scientists often appeal to when investigating which aspects of our minds are the most basic all suggest that it is the capacity to represent knowledge – not belief – that is the more basic component of the theory of mind. Primate study indicates that our ancestors may have begun representing knowledge states before evolving the capacity to think about beliefs. Studies of human infant theory of mind find that infants begin to track what others know before tracking what others believe, and young children can talk about and make predictions using knowledge representations long before belief representations. Tests of the automatic theory of mind in adult humans suggest that representations of knowledge may happen more automatically and effortlessly than representations of beliefs. Also, evidence from patient populations demonstrates that an ability to represent beliefs can be disrupted whereas knowledge attributions remain comparatively preserved.

5. Do attributions of *knowledge* depend on *belief*?

Thus far, we have been reviewing evidence for the existence of a comparatively basic theory of mind representation that shares some of the signature features of knowledge. However, many of the studies on which we've focused have not specifically employed the concepts of *knowledge* and *belief*. Going forward, we will focus more specifically on the explicit representation of the concepts of knowledge and belief.

Much of the relevant studies have been conducted in the field of experimental philosophy. While we review the evidence in detail below, the lesson that comes from this study should sound familiar at this point: knowledge representations seem to be more basic than belief representations, and representations of knowledge do not depend on representations of belief. This conclusion is supported by the fact that response times for knowledge assessments are faster than response times for belief assessments (Section 5.1), that knowledge attributions sometimes occur when belief attributions do not (Section 5.2), that in the best-fitting causal models of the process of mental-state attribution, the ascription of belief does not cause the ascription of knowledge

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

(Section 5.3), and that knowledge attributions are better predictors of behavior than attributions of belief (Section 5.4). We take up each piece of evidence in turn.

5.1 Response times

Consider again the claim that people attribute knowledge by first determining that someone has a belief and then also checking to ensure that this belief has certain further properties (as outlined in Section 3.1). One way of investigating whether this claim is correct is to examine how quickly people are able to make knowledge and belief attributions. If attributing belief is a necessary step in attributing knowledge, then attributions of belief should be faster than attributions of knowledge. Recent study tested this prediction (Phillips et al., 2018).

In one study, participants read about agents who either (a) had a true belief about some proposition *p*, (b) were ignorant and thus had no belief regarding *p*, or (c) believed some other proposition *q* that was inconsistent with *p* (Phillips et al., 2018). After reading about agents in these states, participants were asked whether the agent "knows" that *p*, or instead whether the agent "thinks" that *p*. They were instructed to answer as quickly and accurately as they possibly could.

Participants were systematically faster in both attributing and denying knowledge than in attributing or denying belief – precisely the opposite of the predictions of the belief-as-more-basic view. This pattern was found to extend cross-linguistically, as well, even for a language where the term "think" is *more* frequent than the term "know": French participants were faster to correctly decide what an agent knows ("sait"), than what an agent thinks ("pense"). This provides clear evidence that people's attribution of knowledge cannot depend on a prior attribution of belief.

5.2 Patterns of knowledge attribution versus belief attribution

Still, it might be thought that one prediction that follows from the belief-as-more-basic view is clearly right. Specifically, it should be that all cases in which people are willing to say that someone has knowledge are also cases in which people would be willing to say that someone has the corresponding belief. After all, how could it possibly turn out that a person knows something if she does not even believe it?

Surprisingly, however, results from experimental philosophy provide a reason to doubt that even this prediction is correct. In an important study, researchers tested this claim (Myers-Schulz & Schwitzgebel, 2013; see also Radford, 1966 and Murray, Sytsma, & Livengood, 2012). In one of the scenarios used in this study, participants read about an "unconfident examinee," Kate, who studied very hard for an exam on English history. The exam's final question asked about the date Queen Elizabeth died. Kate had studied this fact many times, but she was not confident that she recalled the answer, so she decided to "just guess" and writes down "1603," which is in fact the correct answer. The vast majority of participants agreed that Kate knew that Queen Elizabeth died in 1603, but only a small minority agreed she believed that. A similar pattern emerged across the other scenarios.

5.3 Does belief attribution predict knowledge attribution?

Thus far, we have been asking whether there are cases in which people are inclined to say that an agent does know something but does not believe it. However, testing whether knowledge attributions are accompanied by belief attributions provides limited information about how these judgments are processed. Even if people strongly tend to attribute knowledge only if they will also attribute belief, it could still be that belief attribution is not central to the psychological process of knowledge attribution.

A series of recent studies suggest that people do *not* base their knowledge attributions on belief attributions. In one set of studies (Turri & Buckwalter, 2017), researchers asked participants to read simple stories about agents and then record several judgments. These judgments included whether the agent knows a particular proposition, along with judgments about several factors that many theorists associate with knowledge, including whether the relevant proposition is true, whether the agent thinks that it is true, whether the agent has good evidence for thinking that it is true, and whether the agent should base a decision on it. In a multiple linear regression model used to predict knowledge attributions, the strongest predictors were judgments about whether the proposition was true and whether the agent should make a decision based on it. Belief attributions did not predict knowledge attributions even when a large number of other relevant factors were controlled (Turri & Buckwalter, 2017). In another set of studies using a similar paradigm, researchers instead used a causal search algorithm to study the relationship among the judgments. In the best-fitting causal model, no kind of belief attribution was found to cause knowledge attributions (Turri, Buckwalter, & Rose, 2016). The upshot of this research is that, even if it turns out that there is some form of belief that is entailed by knowledge, there is currently no evidence that even this minimal form of belief consistently plays a role in the formation of knowledge representations.

5.4 Knowledge and belief in action prediction

A dominant perspective in cognitive science is that our ordinary predictions of others' behavior rely primarily on which beliefs we attribute (e.g., Baker, Saxe, & Tenenbaum, 2009; Rakoczy, 2009; Tomasello, Call, & Hare, 2003). For example, we predict that a traveler will take his umbrella because we attribute to him the belief that it will rain and the desire to stay dry. With the belief and desire attributions in place, there is little additional study left for attributions of knowledge to do in predicting action.

A recent study tested this possibility using a simple paradigm. Participants read brief texts about an agent in various situations which varied the information the agent had access to (e.g., an agent who was following another person and who could or could not see where they turned). After reading the vignette, participants made a belief attribution, a knowledge attribution, and a behavioral prediction. The key question was whether the behavioral prediction would be more strongly predicted by the belief attribution or the knowledge attribution. Knowledge attributions consistently predicted behavioral predictions more strongly than belief attributions did. Moreover, a causal search algorithm suggested that knowledge attributions caused behavioral predictions in contexts where belief attributions did not (Turri, 2016a). Whereas previous research has demonstrated a unique role for knowledge attributions in evaluating how people *should* behave (e.g., Turri, 2015a, 2015b; Turri, Friedman, & Keefner, 2017), these findings suggest a previously unrecognized role for knowledge attributions in predicting how people *will* behave.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

6. Stepping back

The goal of this paper has been to explore the evidence for two competing views of the basic way in which we make sense of others' minds. We saw that every tool used to date to test which kind of representation is more basic – comparative study (Section 4.1), developmental study (Section 4.2), automatic processing in human adults (Section 4.3), and study with patient populations (Section 4.4) – converges on a clear picture: knowledge attribution appears to be a more basic capacity than belief attribution. Moreover, research from experimental philosophy provided independent evidence that knowledge does not depend on belief (Section 5). Critically, we have also illustrated that the theory of mind capacity revealed by these various methods exhibits a set of signature features that are specific to knowledge (Section 2): (i) it is factive, (ii) it is not just true belief, (iii) it allows for egocentric ignorance, and (iv) it is not modality-specific.

6.1. Why knowledge?

A natural question that remains unanswered is why such a capacity for knowledge representation would have ended up being one that is cognitively basic. While there is some good evidence that knowledge representations are used for action prediction (Section 5.4), there are also many instances in which the kind of capacity we provided evidence for will be poorly suited to predicting other agents' behavior. For example, this capacity for knowledge representation would never allow you to predict others' behavior if they happen to disagree with you about the way things are since it only supports factive representations. Also, it is similarly useless when others do agree with you but do so for the wrong reasons – that is, when they have a true belief that falls short of knowledge. This is odd. Understanding *why* someone believes what they do seems entirely unnecessary for predicting their behavior. To do that, you only need to know *what* they believe. Moreover, knowledge representations allow you to represent others as knowing things that you do not know. It is easy to see why this is not particularly useful for action prediction. Imagine a ball was placed in one of two boxes, but you do not know which one. Knowing that someone else knows where the ball is will do you little good in predicting where they will look for it. The question before us is this: given that our more basic ability to represent knowledge has signature features that seem oddly ill-designed for predicting others' behavior, what else might it be for? A promising alternative picture is that the basic capacity for knowledge representation evolved for learning from others.

It is not hard to see how representations of knowledge are well-suited for learning about the extra-mental world (Craig, 1990). Return to the example of a ball being in one of the two boxes, but imagine that you simply want to know where the ball is. From this perspective, the features that are unique to knowledge begin to make perfect sense. For example, you likely do not want your understanding of the ball's location be informed by what someone else *merely believes* about the location of the ball, especially when those beliefs fall short of genuine knowledge – either because they are in conflict with your understanding of the world or because the reason the person came to form them is deviant in some way (Gettier, 1963). In either case, the other person's beliefs will not be a reliable guide to the way the world actually is, and thus if you want to learn about the world, it would be better simply to ignore others' beliefs under such conditions. Moreover, the rather sophisticated capacity to represent others as knowing more than you makes perfect sense here too. While it is clearly not useful for predicting which box the other person will look in, it is incredibly useful for determining who can accurately inform you about where to look. This feature of knowledge (and its contrast with belief) is even reflected in the language we use when talking about others' mental states. We can felicitously talk about others as knowing *where* something is, knowing *how* something is done, or even knowing *who* did something. But we cannot similarly talk about others as believing *where* something is, believing *how* something is done, or believing *who* did something (Egré, 2008; Hintikka, 1975). It is knowledge, but not belief, that allows us to represent others as reliable guides to the actual world.

In short, a promising answer to the question "Why knowledge?" is that knowledge representations are fundamental because they allow us to learn from others about the world. Obviously, this does not mean that we could never *use* knowledge representations to predict others' actions; in fact, we have provided clear evidence that we can and do (Section 5.4). Rather the suggestion is that the capacity for knowledge representation is clearly better designed for learning about the extra-mental world rather than for predicting others' actions, and thus is more likely to have originated for the former, even if it can also be usefully employed for the latter.

6.2. Learning from knowledge

If this hypothesis is correct, the literature on learning from others might offer us valuable clues about the nature of knowledge representations and their pervasive role in cognition. To take one example, the evidence we have reviewed on nonhuman primates suggests that they can use their capacity for knowledge representation to learn, for example, about the location of food based on what others know (Krachun et al., 2009). More generally, a large body of research has demonstrated that nonhuman primates, such as chimpanzees, have an impressive ability to learn from conspecifics, whether in gaining knowledge of how to forage for food (e.g., Rapaport and Brown, 2008) or how to solve novel problems (Call, Carpenter, & Tomasello, 2005; Call & Tomasello, 1995; Myowa-Yamakoshi & Matsuzawa, 2000). Indeed, even capuchin monkeys have an ability to learn foraging techniques from others in a way that is notably sensitive to instances in which others are comparatively more knowledgeable, for example, demonstrations involving novel techniques or unfamiliar foods (Perry, 2011). Obviously, there is good reason to think that this ability for learning from others is relying on representations of others' knowledge rather than their beliefs. Not only are knowledge representations generally better suited for learning about the extra-mental world, but there is little evidence for an ability for belief representation in chimpanzees, and even less in the case of monkeys (Section 4.1). Of course, the point here is not that every instance of learning from others is an instance of knowledge representation – there are all kinds of strategies one can use to learn from others (Heyes, 2016). Our point is just that if you have the capacity for knowledge representation, which we think nonhuman primates do (Section 4.2), then you have a capacity that is well-suited for helping you learn from others.

Similarly, the comparatively basic nature of knowledge representations fits nicely with the literature on learning from others in humans (e.g., Buckwalter & Turri, 2014; Koenig & McMyler, 2019; Mills, 2013; Sobel & Kushnir, 2013). As recently reviewed by Harris et al. (2018), an impressive body of research has documented the capacity to understand others as sources of unknown

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

information from early in human infancy. For example, when young children do not know something themselves, they often request that information from others who know more than they do, even within the first year of life (Kovács et al., 2014), and then selectively learn from others who are knowledgeable rather than not (Hermes, Behne, Bich, Thielert, & Rakoczy, 2018; Moses, Baldwin, Rosicky, & Tidball, 2001). They also seek new knowledge from others who are more likely to understand the relevant part of the world – for example, looking to an experimenter rather than their mother for information about a novel toy (Kim & Kwak, 2011) or attending selectively to information from others who have demonstrated expertise in a given topic (Stenberg, 2013). Moreover, they will specifically ignore information from others who have demonstrated themselves to be unreliable as young as eight-month-old (Begus & Southgate, 2012; Brooker & Poulin-Dubois, 2013; Harris & Corriveau, 2014; Koenig & Harris, 2005; Poulin-Dubois & Brosseau-Liard, 2016; Tummeltshammer, Wu, Sobel, & Kirkham, 2014; Zmyj, Buttelmann, Carpenter, & Daum, 2010). As reviewed above (Section 4.2), the current best evidence suggests that children must be relying on representations of knowledge rather than belief in determining from whom to learn.

From one perspective, this kind of proposal may seem surprising or counterintuitive. From another perspective, however, it seems obvious: when parents teach their children some facts about the world, it does not primarily involve them teaching their children to better understand what they *think* about the world; they are primarily teaching their children to better understand the way the world actually is. Put more simply, we teach others (and expect them to learn) about what we know, not what we believe.

6.2.1. Learning from others, cultural evolution, and what is special about humans

A capacity for reliably learning from others is critically important not only within a single lifespan, but also across them – at the level of human societies. Indeed, this capacity to reliably learn from others has been argued to be essential for human's unique success in the accumulation and transmission of cultural knowledge (e.g., Henrich, 2015; Heyes, 2018). Perhaps unsurprisingly, the argument we have made about the primary role of knowledge representations in cognition fits nicely with this broad view of why humans have been so successful: it is likely supported by our comparatively basic theory of mind representations.

At the same time, this suggestion cuts against another common proposal for which ability underwrites the wide array of ways in which humans have been uniquely successful, namely their ability to represent others' beliefs (Baron-Cohen, 1999; Call & Tomasello, 2008; Pagel, 2012; Povinelli & Preuss, 1995; Tomasello, 1999; Tomasello, Kruger, & Ratner, 1993). While the ability to represent others' beliefs may indeed turn out to be unique to humans and critically important for some purposes, it does not seem to underwrite humans' capacity for the accumulation of cultural knowledge. After all, precisely at the time in human development when the vast majority of critical learning occurs (infancy and early childhood), we find robust evidence for a capacity for knowledge rather than belief representation (Section 4.2).

6.3. A call to arms

Since the 1970s, research has explored belief attribution in a way that brings together numerous areas of cognitive science. Our understanding of belief representation has benefitted from a huge set of interdisciplinary discoveries from developmental studies, cognitive neuroscience, primate cognition, experimental philosophy, and beyond. The result for this empirical ferment has been extraordinary, giving us lots of insight into the nature of belief representation.

We hope this paper serves as a call to arms for cognitive scientists to join researchers who have already begun to do the same for knowledge representation. Our hope is that we can marshal the same set of tools and use them to get a deeper understanding of the nature of knowledge. In doing so, we may gain better insight into the kind of representation that may – at an even more fundamental level – allow us to make sense of others' minds.

Acknowledgments. We would like to thank Jorie Koster-Hale, Joe Henrich, Brent Strickland, Angelo Turri, Evan Westra, and Timothy Williamson.

Financial support. This research was supported in part by funding to JT by the Social Sciences and Humanities Research Council of Canada (SSHRC: 435-2015-0598) and the Canada Research Chairs Program (CRC: 950-231217).

Conflict of interest. None.

Notes

1. This search was conducted using advanced search function on Google Scholar (scholar.google.com). A search for all entries that had an exact match for the term "theory of mind" and additionally had an exact match for either the term "belief task" or "belief test" or "belief condition" yielded 7,930 results. However, a search for all entries that had an exact match for "theory of mind" and additionally had an exact match for either "knowledge task" or "knowledge test" or "ignorance task" or "ignorance test" returned just 897 results. These searches were conducted by the first author on January 12, 2018.
2. Importantly, a control condition showed that primate subjects were able to ignore the competitor's reach when subjects had been directly shown that the food was actually in the other container, suggesting that their performance was not driven by blindly following a reach.

References

Apperly, I. A. (2010). *Mindreaders: The cognitive basis of "theory of mind."* Psychology Press. http://doi.org/10.4324/9780203833926.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a noninferential theory of mind task. *Cognition*, *106*, 1093–1108.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*, 841–844.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, *45*(1), 190–201.

Armstrong, D. M. (1973). *Belief, truth, and knowledge*. Cambridge University Press.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*, 0064.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

Bardi, L., Desmet, C., & Brass, M. (2018). Spontaneous theory of mind is reduced for nonhuman-like agents as compared to human-like agents. *Psychological Research*, *83*, 1571–1580. https://doi.org/10.1007/s00426-018-1000-0.

Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology*, *7*, 113–127. doi: 10.1111/j.2044-835X.1989.tb00793.x.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Baron-Cohen, S. (1999). The evolution of a theory of mind. In M. Corballis & S. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 261–277). Oxford University Press. doi:10.1093/acprof:oso/9780192632593.003.0013.

Baron-Cohen, S., & Goodhart, F. (1994). The 'seeing-leads-to-knowing' deficit in autism: The Pratt and Bryant probe. *British Journal of Developmental Psychology*, *12*(3), 397–401.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201–221.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford University Press.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, *30*, 359–375.

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, *1*(4), 557–560.

Birch, S. A., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, *14*(3), 283–286.

Bloom, L., Rispoli, M., Gartner, B., & Hafitz, J. (1989). Acquisition of complementation. *Journal of Child Language*, *16*(1), 101–120.

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*, 439–448.

Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy*, *18*, E46–E68.

Buckwalter, W., & Turri, J. (2014). Telling, showing and knowing: A unified theory of pedagogical norms. *Analysis*, *74*(1), 16–20. http://doi.org/10.1093/analys/ant092.

Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review*, *125*(3), 409.

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). Great apes distinguish true from false beliefs in an interactive helping task. *PLOS ONE*, *12* (4), e0173793.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606–637.

Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: Chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Animal Cognition*, *8*(3), 151–163.

Call, J., & Santos, L. R. (2012). Understanding other minds. In J. Mitani, P. Kappeler, R. Palombit, J. Call & J. Silk (Eds.), *The evolution of primate societies* (pp. 664–681). University of Chicago Press.

Call, J., & Tomasello, M. (1995). The use of social information in the problem-solving of orangutans (*Pongo pygmaeus*) and human children (*Homo Sapiens*). *Journal of Comparative Psychology*, *109*, 308–320.

Call, J., & Tomasello, M. (1999). A nonverbal theory of mind test. The performance of children and apes. *Child Development*, *70*, 381–395.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, *28*(2), 141–172.

Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162. http://doi.org/10.1007/s13164-015-0259-y.

Carter, J. A., Gordon, E. C., & Jarvis, B. (Eds.). (2017). *Knowledge-First: Approaches in epistemology and mind*. Oxford University Press.

Chisholm, R. M. (1977). *Theory of knowledge* (2d ed.). Prentice-Hall.

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 454.

Craig, E. (1990). *Knowledge and the state of nature: An essay in conceptual synthesis*. Oxford University Press. doi: 10.1093/0198238797.001.0001.

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 958–975. http://dx.doi.org/10.1037/0096-1523.20.5.958.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*(5880), 1217–1220.

Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*(4), 568–570.

Dennett, D. (1989). *The intentional stance*. MIT Press.

de Villiers, J. (1995). Questioning minds and answering machines. In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of the Boston university conference on language development*, *19*, 20–36. Cascadilla Press.

de Villiers, J., & de Villiers, P. (2000). Linguistic determinism and false belief. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Psychology Press.

de Villiers, J., & Pyers, J. (1997). Complementing cognition: The relationship between language and theory of mind. In *Proceedings of the 21st annual Boston University conference on language development* (pp. 136–147). Cascadilla Press.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46*, 12–30. doi: 10.1016/j.cogdev.2018.01.001.

Drayton, L., & Santos, L. R. (2018). What do monkeys know about others' knowledge? *Cognition*, *170*, 201–208.

Drayton, L. A., & Santos, L. R. (2016). A decade of theory of mind research on Cayo Santiago: Insights into rhesus macaque social cognition. *American Journal of Primatology*, *78*(1), 106–116.

Dretske, F. (1981). *Knowledge and the flow of information*. MIT Press.

Dudley, R. (2018). Young children's conceptions of knowledge. *Philosophy Compass*, *13* (6), e12494.

Dudley, R., Orita, N., Hacquard, V., & Lidz, J. (2015). Three-year-olds' understanding of know and think. In *Experimental perspectives on presuppositions* (pp. 241–262). Springer.

Dungan, J., & Saxe, R. (2012). Matched false-belief performance during verbal and nonverbal interference. *Cognitive Science*, *36*, 1148–1156. doi: 10.1111/j.1551-6709.2012.01248.x.

Egré, P. (2008). Question-embedding and activity. *Grazer Philosophische Studien*, *77*(1), 85–125.

El Kaddouri, R., Bardi, L., De Bremaeker, D., Brass, M., & Wiersema, J. R. (2019). Measuring spontaneous mentalizing with a ball detection task: Putting the attention-check hypothesis by Phillips and colleagues (2015) to the test. *Psychological Research*, *84*, 1–9. https://doi.org/10.1007/s00426-019-01181-7.

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.

Fabricius, W. V., & Imbens-Bailey, A. L. (2000). False beliefs about false beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning about the mind* (pp. 267–280). Psychology Press.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Flavell, J. H. (1978). The development of knowledge about visual perception. In C. B. Keasey (Ed.), *The Nebraska symposium on motivation: Vol. 25. Social cognitive development* (pp. 43–76). University of Nebraska Press.

Flavell, J. H. (1992). Perspectives on perspective taking. In H. Beilin & P. B. Pufall (Eds.), *Piaget's theory: Prospects and possibilities* (Vol. 14, pp. 107–139). Erlbaum.

Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, *15*, 447–452.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron*, *32*(6), 969–979.

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 158–163.

Gardner, M. R., Bileviciute, A. P., & Edmonds, C. J. (2018). Implicit mentalising during level-1 visual perspective-taking indicated by dissociation with attention orienting. *Vision*, *2*(1), 3.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Goldman, A. (1967). A causal theory of knowing. *The Journal of Philosophy*, *64*(12), 357–372. doi: 10.2307/2024268.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, *7*(1–2), 145–171.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, *1*(2), 158–171.

Grosse Wiesmann, C., Friederici, A. D., Disla, D., Steinbeis, N., & Singer, T. (2018). Longitudinal evidence for 4-year-olds' but not 2- and 3-year-olds' false belief-related action anticipation. *Cognitive Development*, *46*, 56–68. doi: 10.1016/j.cogdev.2017.08.007.

Hamilton, A. F. D. C., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, *113*(1), 37–44.

Hamlin, K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*, 771–785.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*, 139–151.

Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition*, *101*(3), 495–514.

Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*, *1*(4), 576–77.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Harris, P. L., Bartz, D. T., & Rowe, M. L. (2017a). Young children communicate their ignorance and ask questions. *Proceedings of the National Academy of Sciences*, *114* (30), 7884–7891.

Harris, P. L., & Corriveau, K. H. (2014). Learning from testimony about religion and science. In E. Robinson & S. Einav (Eds.), *Trust and skepticism: Children's selective learning from testimony* (pp. 28–41). Psychology Press.

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273.

Harris, P. L., Ronfard, S., & Bartz, D. (2017b). Young children's developing conception of knowledge and ignorance: Work in progress. *European Journal of Developmental Psychology*, *14*(2), 221–232.

Harris, P. L., Yang, B., & Cui, Y. (2017c). 'I don't know': Children's early talk about knowledge. *Mind & Language*, *32*(3), 283–307.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Hermes, J., Behne, T., Bich, A. E., Thielert, C., & Rakoczy, H. (2018). Children's selective trust decisions: Rational competence and limiting performance factors. *Developmental Science*, *21*(2), e12527.

Heyes, C. (2014a). False belief in infancy: A fresh look. *Developmental Science*, *17*(5), 647–659.

Heyes, C. (2014b). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131–143.

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences*, *20*(3), 204–213.

Heyes, C. (2017). Apes submentalise. *Trends in Cognitive Sciences*, *21*(1), 1–2.

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Hintikka, J. (1975). Different constructions in terms of the basic epistemological verbs: A survey of some problems and proposals. In *The intensions of intentionality and other new models for modalities* (pp. 1–25). D. Reidel.

Hobson, R. P. (1984). Early childhood autism and the question of egocentrism. *Journal of Autism and Developmental Disorders*, *14*(1), 85–104.

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*, 72–80.

Ichikawa, J. J., & Steup, M. (2017). The analysis of knowledge. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy (Fall 2017 Edition)*. Stanford Encyclopedia of Philosophy. Retrieved August 20, 2017, from https://plato.stanford.edu/archives/fall2017/entries/knowledge-analysis.

Johnson, C. N., & Maratsos, M. P. (1977). Early comprehension of mental verbs: Think and know. *Child Development*, *48*, 1743–1747.

Jordan, K. E., & Brannon, E. M. (2006). A common representational system governed by Weber's law: Nonverbal numerical similarity judgments in 6-year-olds and Rhesus Macaques. *Journal of Experimental Child Psychology*, *95*(3), 215–229.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*, 224–234.

Kammermeier, M., & Paulus, M. (2018). Do action-based tasks evidence false-belief understanding in young children? *Cognitive Development*, *46*, 31–39. doi: 10.1016/j.cogdev.2017.11.004.

Kano, F., Krupenye, C., Hirata, S., Call, J., & Tomasello, M. (2017). Submentalizing cannot explain belief-based action anticipation in apes. *Trends in Cognitive Sciences*, *21* (9), 633–634. doi: 10.1016/j.tics.2017.06.011.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences*, *116*(42), 20904–20909.

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015). The goggles experiment: Can chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour*, *105*, 211–221.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*, 3–44.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498–525. doi: 10.2307/1418556.

Kim, G., & Kwak, K. (2011). Uncertainty matters: Impact of stimulus ambiguity on infant social referencing. *Infant and Child Development*, *20*(5), 449–463.

Kiparksy, P., & Kiparksy, C. (1970). Fact. In M. Bierwisch & K. Heidolph (Eds.), *Progress in linguistics: A collection of papers* (pp. 143–173). Mouton & Co. N.V. Publishers.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277.

Koenig, M. A., & McMyler, B. (2019). Testimonial knowledge: Understanding the evidential, uncovering the interpersonal. In M. Fricker, P. Graham, D. Henderson, N. Pederson, J. Wyatt (Eds.), *The Routledge handbook of social epistemology* (pp. 103–114). Routledge.

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy*, *19* (6), 543–557.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibly to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2017). A test of the submentalizing hypothesis: Apes' performance in a false belief task inanimate control. *Communicative & Integrative Biology*, *10*(4), e1343771.

Kulke, L., Johannsen, J., & Rakoczy, H. (2019). Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PloS One*, *14*(3), e0213772–e0213772. https://doi.org/10.1371/journal.pone.0213772.

Kulke, L., Wübker, M., & Rakoczy, H. (2018). Is implicit theory of mind real but hard to detect? Testing adults with different stimulus materials. *Royal Society Open Science*, *6* (7), 190068.

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, *15*(1), 77–95.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*(12), 528–533.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, *6*, 315–324. doi: 10.1111/j.2044-835X.1988.tb01104.x.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, *74*(4), 549–567. doi: 10.1080/00048409612347521.

Lewis, S., Hacquard, V., & Lidz, J. (2012). The semantics and pragmatics of belief reports in preschoolers. *Semantics and Linguistic Theory*, *22*, 247–267.

Locke, J. (1689/1975). In P. H. Nidditch (Ed.), *An essay concerning human understanding*. Oxford University Press.

Low, J., & Edwards, K. (2018). The curious case of adults' interpretations of violation-of-expectation false belief scenarios. *Cognitive Development*, *46*, 86–96.

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*(3), 305–311.

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition*, *121* (3), 289–298.

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149.

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., … Hashimoto, T. (2017). Gettier across cultures. *Noûs*, *51*(3), 645–664.

MacLean, E. L., & Hare, B. (2012). Bonobos and chimpanzees infer the target of another's attention. *Animal Behaviour*, *83*(2), 345–353.

Mar, R. A., Tackett, J. L., & Moore, C. (2010). Exposure to media and theory-of-mind development in preschoolers. *Cognitive Development*, *25*(1), 69–78.

Marshall, J., Gollwitzer, A., & Santos, L. R. (2018). Does altercentric interference rely on mentalizing?: Results from two level-1 perspective-taking tasks. *PLOS ONE*, *13*(3), e0194101.

Marticorena, D. C. W., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science*, *14*(6), 1406–1416. doi: 10.1111/j.1467-7687.2011.01085.x.

Martin, A. (2019). Belief representation in great apes. *Trends in Cognitive Sciences*, *23*(12), 985–986.

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12 months. *Cognition*, *123*(1), 50–60.

Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition*, *130*(3), 300–308.

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375–382.

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*(2), 154.

Meristo, M., & Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition*, *129*(1), 102–113.

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, *49*(3), 404.

Moll, H., Carpenter, M., & Tomasello, M. (2014). Two-and 3-year-olds know what others have and have not heard. *Journal of Cognition and Development*, *15*(1), 12–21.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, *60*(1), 167–171.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences*, *108*(7), 2688–2692.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*(3), 718–735.

Murray, D., Sytsma, J., & Livengood, J. (2012). God knows (but does God believe?). *Philosophical Studies*, *166*(1), 83–107. doi: 10.1007/s11098-012-0022-5.

Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing that *p* without believing that P. *Noûs*, *47*(2), 371–384.

Myowa-Yamakoshi, M., & Matsuzawa, T. (2000). Imitation of intentional manipulatory actions in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *114*, 381–391.

Nagel, J. (2013). Knowledge as a mental state. *Oxford Studies in Epistemology*, *4*, 275–310. doi: 10.1093/acprof:oso/9780199672707.003.0010.

Nagel, J. (2017). Factive and nonfactive mental state attribution. *Mind & Language*, *32*(5), 525–544.

Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford University Press.

Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.

O'Connell, S., & Dunbar, R. (2003). A test for comprehension of false belief in chimpanzees. *Evolution and Cognition*, *9*(2), 131–140.

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

O'Neill, D. K., Astington, J. W., & Flavell, J. H. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development*, *63* (2), 474–490.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.

Pagel, M. (2012). *Wired for culture: Origins of the human social mind*. WW Norton & Company.

Papafragou, A., Li, P., Choi, Y., & Han, C. H. (2007). Evidentiality in language and cognition. *Cognition*, *103*(2), 253–299.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, *60*, 689–700.

Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition*, *145*, 77–88. http://dx.doi.org/10.1016/j.cognition.2015.08.006.

Perry, S. (2011). Social traditions and social learning in capuchin monkeys (Cebus). *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366* (1567), 988–996.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development*, *76*(2), 502–517.

Phillips, J., & George, B. R. (2018). Knowledge wh and false beliefs: Experimental investigations. *Journal of Semantics*, *35*(3), 467–494.

Phillips, J., & Norby, A. (2019). Factive theory of mind. *Mind & Language*, *36*, 3–26. https://doi.org/10.1111/mila.12267.

Phillips, J., Ong, D. C., Surtees, A. D., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A second look at automatic theory of mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*, *26*(9), 1353–1367.

Phillips, J., Strickland, B., Dungan, J., Armary, P., Knobe, J., & Cushman, F. (2018). *Evidence for evaluations of knowledge prior to belief*. Proceedings of the Fortieth Annual Conference of the Cognitive Science Society.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology*, *47*(1), 116–129.

Plato. (380BCE/1997). The republic. In J. M. Cooper (Ed.), G. M. A. Grube & C. D. C. Reeve (Trans.), *Plato: Complete works*. Hackett.

Poulin-Dubois, D., & Brosseau-Liard, P. (2016). The developmental origins of selective social learning. *Current Directions in Psychological Science*, *25*(1), 60–64.

Povinelli, D. J., & Preuss, T. M. (1995). Theory of mind: Evolutionary history of a cognitive specialization. *Trends in Neurosciences*, *18*(9), 418–424.

Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, *46*, 40–50. doi: 10.1016/j.cogdev.2017.10.004.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*(4), 973–982.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Priewasser, B., Rafetseder, E., Gargitter, C., & Perner, J. (2018). Helping as an early indicator of a theory of mind: Mentalism or teleology? *Cognitive Development*, *46*, 69–78.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.

Radford, C. (1966). Knowledge – by examples. *Analysis*, *27*(1), 1–11.

Rakoczy, H. (2009). Executive function and the development of belief-desire psychology. *Developmental Science*, *13*(4), 648–661. doi: 10.1111/j.1467-7687.2009.00922.x.

Rapaport, L. G., & Brown, G. R. (2008). Social influences on foraging behavior in young nonhuman primates: Learning what, where, and how to eat. *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(4), 189–201.

Reed, T., & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. *Journal of Autism and Developmental Disorders*, *20*(4), 555–567.

Rose, D. (2015). Belief is prior to knowledge. *Episteme; Rivista Critica Di Storia Delle Scienze Mediche E Biologiche*, *12*(3), 385–399.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–66. doi: 10.1037/a0018729.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929–937. doi: 10.1037/a0035175.

Santos, L. R., Barnes, J. L., & Mahajan, N. (2005). Expectations about numerical events in four lemur species (*Eulemur fulvus*, *Eulemur mongoz*, *Lemur catta* and *Varecia rubra*). *Animal Cognition*, *8*(4), 253–262.

Santos, L. R., Nissen, A. G., & Ferrugia, J. (2006). Rhesus monkeys (*Macaca mulatta*) know what others can and cannot hear. *Animal Behaviour*, *71*, 1175–1181.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *Neuroimage*, *19*(4), 1835–1842.

Schmelz, M., Call, J., & Tomasello, M. (2011). Chimpanzees know that others make inferences. *Proceedings of the National Academy of Sciences*, *108*(7), 3077–3079.

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, *23*, 842–847. https://doi.org/10.1177/0956797612439070.

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, *129*(2), 410–417.

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society Open Science*, *5*(5), 172273.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883–885. http://dx.doi.org/10.1126/science.1176170.

Shahaeian, A., Nielsen, M., Peterson, C. C., & Slaughter, V. (2014). Cultural and family influences on children's theory of mind development: A comparison of Australian and Iranian school-age children. *Journal of Cross-Cultural Psychology*, *45*(4), 555–568.

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*(5), 1239.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, *14*(3), 301–321.

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120* (4), 779.

Sodian, B., Thoermer, C., & Dietrich, N. (2006). Two-to four-year-old children's differentiation of knowing and guessing in a non-verbal task. *European Journal of Developmental Psychology*, *3*(3), 222–237.

Sosa, E. (1999). How to defeat opposition to Moore. *Noûs*, *33*(13), 141–153. doi: 10.1111/0029-4624.33.s13.7.

Sowalsky, E., Hacquard, V., & Roeper, T. (2009). Is PP opacity on the path to false belief. *Generative Approaches to Language Acquisition North America* (*GALANA*), *3*, 263–261.

Spelke E. S. (2004). Core knowledge. In N. Kanwisher & J. Duncan (Eds.), *Attention and performance: Functional neuroimaging of visual cognition* (Vol. *20*, pp. 29–56). Oxford University Press.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, *124* (3), 272–283.

Stenberg, G. (2013). Do 12-month-old infants trust a competent adult? *Infancy*, *18*(5), 873–904.

Stich, S. (2013). Do different groups have different epistemic intuitions? A reply to Jennifer Nagel. *Philosophy and Phenomenological Research*, *87*(1), 151–178.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586.

Surtees, A., Apperly, I., & Samson, D. (2016a). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52.

Surtees, A., Samson, D., & Apperly, I. (2016b). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Surtees, A. D., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, *83*(2), 452–460.
Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, *30*, 75–86. doi: 10.1111/j.2044-835X.2011.02063.x.
Tahiroglu, D., Moses, L. J., Carlson, S. M., Mahy, C. E., Olofson, E. L., & Sabbagh, M. A. (2014). The children's social understanding scale: Construction and validation of a parent-report measure for assessing individual differences in children's theories of mind. *Developmental Psychology*, *50*(11), 2485–2497.
Tan, J., & Harris, P. L. (1991). Autistic children understand seeing and wanting. *Development and Psychopathology*, *3*(2), 163–174.
Tardif, T., & Wellman, H. M. (2000). Acquisition of mental state language in Mandarin-and Cantonese-speaking children. *Developmental Psychology*, *36* (1), 25.
Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.
Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, *115*(34), 8491–8498.
Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Sciences*, *7*(4), 153–156.
Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495–552.
Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, *15*(4), 434–444.
Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101* (1), 80–102. http://dx.doi.org/10.1037/0033-295X.101.1.80.
Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science*, *25*(9), 1730–1738.
Turri, J. (2015a). Evidence of factive norms of belief and decision. *Synthese*, *192*, 4009–4030. doi: 10.1007/s11229-015-0727-z.
Turri, J. (2015b). Knowledge and the norm of assertion: A simple test. *Synthese*, *192*(2), 385–392. doi: 10.1007/s11229-014-0573-4.
Turri, J. (2016a). Knowledge attributions and behavioral predictions. *Cognitive Science*, *41* (8), 2253–2261.
Turri, J., & Buckwalter, W. (2017). Descartes's schism, Locke's reunion: Completing the pragmatic turn in epistemology. *American Philosophical Quarterly*, *54*(1), 25–46.
Turri, J., Buckwalter, W., & Rose, D. (2016). Actionability judgments cause knowledge judgments. *Thought: A Journal of Philosophy*, *5*(3), 212–222. http://doi.org/10.1002/tht3.213.
Turri, J., Friedman, O., & Keefner, A. (2017). Knowledge central: A central role for knowledge attributions in social evaluations. *Quarterly Journal of Experimental Psychology*, *70*(3), 504–515.
van der Wel, R. P., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128–133.
Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, *17*(6), 872–879.
Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of- mind development: The truth about false belief. *Child Development*, *7*, 655–684.
Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, *17*(12), 1075–1081.
Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.
Whiten, A. (2013). Humans are not alone in computing how others see the world. *Animal Behaviour*, *86*(2), 213–221.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press. doi:10.1093/019925656X.001.0001.
Woolley, J. D., & Wellman, H. M. (1993). Origin and truth: Young children's understanding of imaginary mental representations. *Child Development*, *64*(1), 1–17.
Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11.
Yirmiya, N., Sigman, M., & Zacks, D. (1994). Perceptual perspective-taking and seriation abilities in high-functioning children with autism. *Developmental Psychopathology*, *6*, 263–272. doi: 10.1017/S0954579400004570.
Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, *106*(4), 208–220.

Open Peer Commentary

Beyond knowledge versus belief: The contents of mental-state representations and their underlying computations

Mika Asaba, Aaron Chuey, and Hyowon Gweon

Department of Psychology, Stanford University, Stanford, CA 94305, USA.
masaba@stanford.edu; chuey@stanford.edu; hyo@stanford.edu
https://sll.stanford.edu

doi:10.1017/S0140525X20001879, e141

Abstract

Moving beyond distinguishing knowledge and beliefs, we propose two lines of inquiry for the next generation of theory of mind (ToM) research: (1) characterizing the contents of different mental-state representations and (2) formalizing the computations that generate such contents. Studying how children reason about what others think of the self provides an illuminating window into the richness and flexibility of human social cognition.

We agree with Philips et al. that examining a greater variety of epistemic states will enrich our understanding of the origins and development of mentalizing capacities, and appreciate their distinction between knowledge and belief. Importantly, however, categorizing these mental states and asking at what age children can represent them are just the first steps toward characterizing the richness and flexibility of our social-cognitive capacities. Looking back on over a decade of research attempting to identify the earliest signatures of belief attribution (e.g., Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007; but see also Dörrenberg, Rakoczy, & Liszkowski, 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018), we caution against yet another arms race to determine which representation is "more basic" or "more critical" to social learning. Instead, as cognitive scientists, we find ourselves asking: *How* are these mental-state representations cognitively distinct from one another, and what cognitive mechanisms support these representations?

Imagine your roommate is watching as you try to open a jar that just won't budge. You might say that your roommate *knows* you failed to open the jar, *believes* the jar is difficult to open, or even *thinks* that you are too weak to open it. Although you would be comfortable using these words – know, believe, think – to describe your roommate's mental states, their contents differ from what is typically studied in the theory of mind (ToM) literature. Rather than reflecting verifiable external states of the world (e.g., the location of the jar), these mental states concern outcomes of intentional, goal-directed actions (e.g., failure to open a jar) and outputs of additional inferences based on observed action–outcome relationships, such as subjective evaluations about abstract qualities of objects (e.g., the jar may be difficult to open for some but not others) and agents, including oneself (e.g., you may be weak compared to some but not others).

To understand how our mind flexibly generates, represents, and attributes these mental states, we need to move beyond traditional concepts and empirical paradigms that have dominated the past

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

few decades of research on ToM. To this end, young children's reasoning about how others represent and infer abstract qualities of people (including the self) can provide a particularly illuminating window into the richness and flexibility of human social cognition. Our recent study finds that children, by 4 years of age, are already capable of reasoning about what others think of them after observing their own failures or successes (Asaba & Gweon, 2018; Asaba & Gweon, 2021). Looking forward, we propose focusing our efforts on two related lines of inquiry: (1) characterizing the *contents* of mental states that children (and nonhuman primates) can represent and (2) understanding and formalizing the *computations* (i.e., inferential processes) that give rise to such representations.

1. Contents of mental states

If your roommate observed your failure to open the jar countless times every day, you may intuitively feel that your roommate "knows" you cannot open the jar. Similarly, we might use the word "know" to describe one's various attitudes toward someone else (e.g., "Sally knows that Ann is generous, funny, and competent"), especially when we suspect one has strong evidence about these qualities. However, these are inherently subjective evaluations that do not have objective, verifiable criteria for determining their truth value; they can only be expressed as the degree to which one "believes" X is true, rather than as a Boolean value (i.e., either true or false). Although people often describe these mental states using knowledge-laden language, these contents go beyond the scope of what Philips et al. would consider as knowledge.

Critically however, the content of these representations also differ from the content of beliefs that are typically studied in the ToM literature. Rather than observable states of the physical world that are verifiable via perception (e.g., "Sally knows her toy is in the box" or "Sally thinks her toy is in the box"), these belief states concern abstract properties of agents that must be inferred from an agent's behaviors or other social sources of information (e.g., others' evaluative feedback or testimony, such as "Ann is very generous"). Beyond distinguishing knowledge versus belief, we need more research on *why* children find some mental-state contents harder to attribute than others.

2. Underlying computations

Relatedly, the process by which we attribute mental states about internal qualities of agents may involve more complex computations than those concerning external states of the world. When an agent has direct perceptual access to a world-state, there is a one-to-one correspondence between what they see and what they represent. When the agent loses perceptual access while the world-state changes, the agent is rendered ignorant (i.e., Sally does not know where her toy is) or mistaken (i.e., Sally falsely believes that her toy is in the box). Although understanding the relationship between an agent's perception and their resulting epistemic state is already an impressive feat, representing others' beliefs about internal qualities is even more so; these representations cannot be derived from perceptual access alone, and require further inferences based on an intuitive understanding of how observations of an agent's goal-directed actions give rise to representations of the agent's abstract qualities.

Much of the prior literature on ToM development has studied how children represent others' ignorance or false beliefs that are decoupled from reality. However, these representations reflect only a fraction of the mental states we encounter in our everyday social interactions. When Sally observes Ann bake a delicious cake, get a "D" on a math exam, or donate $20 dollars to charity, what kind of mental states might children attribute to Sally? These representations could be about anyone, but they are especially powerful when they concern qualities of the self: Does Sally think I am good at baking? Terrible at math? Generous or stingy? Although we, as adults, naturally entertain these thoughts, more research is needed to understand how young children integrate their understanding of the physical and social world to attribute these nuanced mental states. Recent computational research has made major advances in formalizing the generative process by which an agent's observation gives rise to beliefs about the external world (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017; Jara-Ettinger, Schulz, & Tenenbaum, 2020); these approaches can provide important insights here as well.

Ultimately, the ability to represent others' mental states is both a blessing and a curse. When directed at us, it inspires our motivation to learn and responsibility for our own actions; when gone awry, it dogs us with unnecessary worries about how others might evaluate us. Yet, for better or worse, we rely on these abilities to learn from others in a complex social world filled with competition, cooperation, and collaboration. The richness of human cultural knowledge comes from our ability to appreciate, evaluate, criticize, and communicate a host of abstract thoughts. By studying how the human mind supports these rich mental-state inferences, we can better understand how humans harness these capacities to learn from others and help others learn.

References

Asaba, M., & Gweon, H. (2018). Look, I can do it! Young children forego opportunities to teach others to demonstrate their own competence. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp. 106–111). Cognitive Science Society.

Asaba, M., & Gweon, H. (2021). Young children rationally revise and maintain what others think of them. *PsyArxiv*. Preprint doi: 10.31234/osf.io/yxhv5.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*, 1–10. doi: 10.1038/s41562-017-0064.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46*, 12–30. doi: 10.1016/j.cogdev.2018.01.001.

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naïve utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334. doi: 10.1016/j.cogpsych.2020.101334.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258. doi: 10.1126/science.1107621.

Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, *46*, 40–50. doi: 10.1016/j.cogdev.2017.10.004.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586. doi: 10.1111/j.1467-9280.2007.01943.x.

Infants actively seek and transmit knowledge via communication

Marina Bazhydai[a] and Paul L. Harris[b]

[a]Psychology Department, Fylde College, Lancaster University, Lancaster LA1 4YW, UK and [b]Harvard Graduate School of Education, Cambridge, MA 02138, USA.
m.bazhydai@lancaster.ac.uk
paul_harris@gse.harvard.edu

doi:10.1017/S0140525X20001405, e142

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Abstract

Supporting the central claim that knowledge representation is more basic than belief representation, we focus on the emerging evidence for preverbal infants' active and selective communication based on their representation of both knowledge and ignorance. We highlight infants' ontogenetically early deliberate information seeking and information transmission in the context of active social learning, arguing that these capacities are unique to humans.

Supporting the central claim of Phillips et al. that knowledge rather than belief representation constitutes the more basic cognitive capacity, we highlight emerging evidence from developmental research on infants' communicative use of such knowledge representations. This commentary suggests that infants actively and selectively seek epistemic input from more knowledgeable others and, in turn, transmit such information to less knowledgeable others. As active participants in the bidirectional exchange of knowledge, infants take an interrogative stance and also themselves act as informants, using developmentally available tools from their preverbal communicative repertoire.

To actively solicit information, infants have been shown to socially reference adults who were more likely to help them resolve an epistemically uncertain situation (Bazhydai, Westermann, & Parise, 2020c; Goupil, Romand-Monnier, & Kouider, 2016; Harris, Bartz, & Rowe, 2017; Stenberg, 2013; Vaish, Demir, & Baldwin, 2011) and point to objects they want to learn about in the presence of a knowledgeable rather than an uninformed person (Begus & Southgate, 2012; Kovács, Tauzin, Téglás, Gergely, & Csibra, 2014; Lucca & Wilbourn, 2018). To actively transmit information in situations where infants themselves were more knowledgeable than their social partners, they have been shown to use informative pointing (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Liszkowski, Carpenter, & Tomasello, 2008; Meng & Hashiya, 2014; O'Neill, 1996) and deliberate action demonstration (Bazhydai, Silverstein, Parise, & Westermann, 2020a; Flynn, 2008; Vredenburgh, Kushnir, & Casasola, 2015) as communicative tools indicative of early emerging, proto-teaching strategies (Strauss & Ziv, 2012).

Not only do infants represent others' states of knowledge versus ignorance, but they also form epistemic expectations and actively seek explanations or clarifications when adults do not act in accordance with those prior expectations (Harris, Koenig, Corriveau, & Jaswal, 2018). For instance, infants expect to learn from previously knowledgeable informants (Begus, Gliga, & Southgate, 2016) and look longer toward adults who provide inaccurate labels for familiar objects (Koenig & Echols, 2003) or object location information incongruent with their true knowledge (Galazka, Gredebäck, & Ganea, 2016). Such enhanced attention to a speaker can be plausibly interpreted as indexing a violation of the expectation that social partners are, by default, reliable rather than misleading in their information provision (Sperber et al., 2010). Furthermore, infants are less likely to subsequently learn from previously untrustworthy informants (Brooker & Poulin-Dubois, 2013; Koenig & Woodward, 2010) or from those who provide information incongruent with what was asked of them (Begus, Gliga, & Southgate, 2014).

These early behaviors indicate that in social situations of epistemic uncertainty, infants act to close both intra- and inter-individual knowledge gaps, ultimately achieving an equal distribution of knowledge upon its social transfer (Harris, 2017; Strauss & Ziv, 2012). Prominent theories of epistemic curiosity (information gap and learning progress; for a review, see Bazhydai, Twomey, & Westermann, 2020b) conclude that the information being sought is inherently factual, as it is in the case of curiosity in social learning (Begus & Southgate, 2018; Harris, 2020). When providing information, evidence to date shows that infants transmit factual information (e.g., where a hidden object is located or how to make a new toy play music).

Notably, and in accord with the proposals by Phillips et al., the ability to actively exchange knowledge does not presuppose a full-blown mentalizing ability: Although infants exert control over their information seeking and information provision, their behaviors are likely proto-metacognitive (but see e.g., Goupil & Kouider, 2016, for evidence of metacognitive sensitivity in infancy) (Harris, 2020; Heyes, 2016; Strauss & Ziv, 2012). For example, when infants indicate what they know to ignorant others, there is no evidence so far to suggest they realize that they possess unique transferable knowledge, or that they deliberately reason about the nature of their communicative behaviors, which nevertheless perform an informative function, as a result of which others can also know. Thus, the intra- and inter-individual epistemic gaps do not have to be realized as such for active social learning to occur.

Crucially, and dissenting from the picture painted by Phillips et al., we argue that this deliberate process of "asking for" knowledge and spontaneously taking steps to pass it on is a distinctively human ability. Although nonhuman animals represent others' knowledge and act in accordance with those representations, in contrast to infants (Harris & Lane, 2014; Ronfard & Harris, 2015), we see little evidence of active information solicitation in them. Similarly, evidence of information transmission remains limited and less diverse and flexible than that of humans (Burdett, Dean, & Ronfard, 2017; Strauss & Ziv, 2012; but see Musgrave et al., 2020, for new evidence of teaching-like behaviors in wild chimpanzees). For example, infants have the capacity to exchange cultural information (e.g., the label or function of an artifact) as opposed to exclusively functional information (e.g., the location of a food source), thereby distinguishing human infants' information exchange from that of any other nonhuman primate. Thus, among the various social-learning strategies that involve the transmission of knowledge from one social partner to another, and which we share to a large extent with nonhuman animals (imitation, emulation, and observation), the active and selective seeking and provision of information among conspecifics appears to be unique to humans.

We are excited about the directions outlined in the paper's call to arms and emphasize the need to investigate the developmental foundations of information seeking and transmission. In light of the likely connections between curiosity and teaching in cultural evolution (van Schaik, Pradhan, & Tennie, 2019), future studies should investigate how the active solicitation of information impacts its subsequent transmission, examining whether the motivation to seek knowledge rather than belief makes information sharing more likely. If knowledge representations are primary, knowledge- rather than belief-based information would be more likely to be both sought and further propagated.

In summary, we support the knowledge-as-more-basic view and propose to strengthen the account by adding to the list of signature properties of knowledge representation, the ability of infants to engage in active social learning as manifested in information seeking and information transmission. These emerging

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

findings support the proposal that knowledge representation as a basic capacity may be shared with other evolutionarily close species, whereas the active communication of knowledge evolved in humans to optimize learning from others and informing others.

Financial support. This study received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Bazhydai, M., Silverstein, P., Parise, E., & Westermann, G. (2020a). Two-year old children preferentially transmit simple actions but not pedagogically demonstrated actions. *Developmental Science*, *23*(5), e12941. doi: 10.1111/desc.12941.

Bazhydai, M., Twomey, K., & Westermann, G. (2020b). Curiosity and exploration. In: Benson, J.B. (Ed.), *Encyclopedia of infant and early childhood development* (Vol. *1*, 2nd ed., pp. 370–378). Elsevier. doi: 10.1016/B978-0-12-809324-5.05804-1.

Bazhydai, M., Westermann, G., & Parise, E. (2020c). "I don't know but I know who to ask": 12-month-olds actively seek information from knowledgeable adults. *Developmental Science*, *23*(5), e12938. doi: 10.1111/desc.12938.

Begus K., Gliga T., & Southgate V. (2014). Infants learn what they want to learn: Responding to infant pointing leads to superior learning. *PLOS ONE 9*(10):e108817. doi: 10.1371/journal.pone.0108817.

Begus, K., Gliga, T., & Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *PNAS: Proceedings of the National Academy of Sciences*, *113*(44), 12397–12402. doi: 10.1073/pnas.1603261113.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617. doi: 10.1111/j.1467-7687.2012.01160.x.

Begus, K., & Southgate, V. (2018). Curious learners: How infants' motivation to learn shapes and is shaped by infants' interactions with the social world. In M. Saylor and P. Ganea (Eds.), *Active learning from infancy to childhood* (pp. 13–37). Springer. doi: 10.1007/978-3-319-77182-3_2.

Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy*, *18*, E46–E68. doi: 10.1111/infa.12027.

Burdett, E. R., Dean, L. G., & Ronfard, S. (2017). A diverse and flexible teaching toolkit facilitates the human capacity for cumulative culture. *Review of Philosophy and Psychology*, *9*(4), 807–818. doi: 10.1007/s13164-017-0345-4.

Flynn, E. (2008). Investigating children as cultural magnets: Do young children transmit redundant information along diffusion chains? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3541–3551. doi: 10.1098/rstb.2008.0136.

Galazka, M. A., Gredebäck, G., & Ganea, P. A. (2016). Mapping language to the mind: Toddlers' online processing of language as a reflection of speaker's knowledge and ignorance. *Cognitive Development*, *40*, 1–8. doi: 10.1016/j.cogdev.2016.07.003.

Goupil, L., & Kouider, S. (2016). Behavioral and neural indices of metacognitive sensitivity in preverbal infants. *Current Biology*, *26*(22), 3038–3045. doi: 10.1016/j.cub.2016.09.004.

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, *113*(13), 3492–3496. doi: 10.1073/pnas.1515129113.

Harris, P. L. (2017). Tell, ask, repair: Early responding to discordant reality. *Motivation Science*, *3*, 275–286.

Harris, P. L. (2020). The point, the shrug, and the question of clarification. In L. Butler, S. Ronfard & K. Corriveau (Eds.), *The questioning child: Insights from psychology and education* (pp. 29–50). Cambridge University Press.

Harris, P. L., Bartz, D. T., & Rowe, M. L. (2017). Young children communicate their ignorance and ask questions. *Proceedings of the National Academy of Sciences*, *114*(30), 7884–7891. doi: 10.1073/pnas.1620745114.

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273. doi: 10.1146/annurev-psych-122216-011710.

Harris, P. L., & Lane, J. D. (2014). Infants understand how testimony works. *Topoi*: *An International Review of Philosophy*, *33*, 443–458. doi: 10.1007/s11245-013-9180-0.

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences*, *20*(3), 204–213.

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, *87*(3), 179–208. doi: 10.1016/S0010-0277(03)00002-7.

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology*, *46*(4), 815–826. doi: 10.1037/a0019664.

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy*, *19*(6), 543–557. doi: 10.1111/infa.12060.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12-and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, *7*(2), 173–187. doi: 10.1207/s15327647jcd0702_2.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108* (3), 732–739. doi: 10.1016/j.cognition.2008.06.013.

Lucca, K., & Wilbourn, M. P. (2018). Communicating to learn: Infants' pointing gestures result in optimal learning. *Child Development*, *89*(3), 941–960. doi: 10.1111/cdev.12707.

Meng, X, & Hashiya, K. (2014). Pointing behavior in infants reflects the communication partner's attentional and knowledge states: A possible case of spontaneous informing. *PLOS ONE*, *9*(9), e107579. doi: 10.1371/journal.pone.0107579.

Musgrave, S., Lonsdorf, E., Morgan, D., Prestipino, M., Bernstein-Kurtycz, L., Mundry, R., & Sanz, C. (2020). Teaching varies with task complexity in wild chimpanzees. *Proceedings of the National Academy of Sciences*, *117*(2), 969–976. doi: 10.1073/pnas.1907476116.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*(2), 659–677. doi: 10.1111/j.1467-8624.1996.tb01758.x.

Ronfard, S., & Harris, P. L. (2015). The active role played by human learners is key to understanding the efficacy of teaching in humans. *Behavioral and Brain Sciences*, *38*, 43–44. doi: 10.1017/S0140525X14000594.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393. doi: 10.1111/j.1468-0017.2010.01394.x.

Stenberg, G. (2013). Do 12-month-old infants trust a competent adult? *Infancy*, *18*(5), 873–904. doi: 10.1111/infa.12011.

Strauss, S., & Ziv, M. (2012). Teaching is a natural cognitive ability for humans. Mind, *Brain, and Education*, *6*(4), 186–196. doi: 10.1111/j.1751-228X.2012.01156.x.

Vaish, A., Demir, O., & Baldwin, D. (2011). Thirteen- and 18-month-old infants recognize when they need referential information. *Social Development*, *20*(3), 431–449. doi: 10.1111/j.1467-9507.2010.00601.x.

van Schaik, C. P., Pradhan, G. R., & Tennie, C. (2019). Teaching and curiosity: Sequential drivers of cumulative cultural evolution in the hominin lineage. *Behavioral Ecology and Sociobiology*, *73*(1), 2. doi: 10.1007/s00265-018-2610-7.

Vredenburgh, C., Kushnir, T., & Casasola, M. (2015). Pedagogical cues encourage toddlers' transmission of recently demonstrated functions to unfamiliar adults. *Developmental Science*, *18*(4), 645–654. doi: 10.1111/desc.12233.

Knowledge *is* belief – and shaped by culture

Andrea Bender[a] and John B. Gatewood[b]

[a]Department of Psychosocial Science & SFF Centre for Early Sapiens Behaviour (SapienCE), University of Bergen, N-5020 Bergen, Norway and [b]Department of Sociology and Anthropology, Lehigh University, Bethlehem, PA 18015-3169, USA.

andrea.bender@uib.no; jbg1@lehigh.edu

https://www.uib.no/en/persons/Andrea.Bender;

https://socanthro.cas.lehigh.edu/content/john-b-gatewood

doi:10.1017/S0140525X20001582, e143

Abstract

Phillips and colleagues claim that the representation of knowledge is more basic than the representation of belief, presupposing them to be categorically distinct mental states with distinct evolutionary purposes. We argue that the relationship between the two is much more complex, is further shaped by culture and language, and leaves its mark on manifestations of theory of mind and teaching.

> "Any act of factual knowing presupposes somebody who believes he knows what is being believed to be known. This person is taking a risk in asserting something, at least tacitly, about something believed to be real outside himself."
>
> — Polanyi (1958, p. 313)

In asking whether the representation of knowledge is more basic than the representation of belief, Phillips and colleagues presuppose that knowledge and belief are categorically distinct mental states. The authors also claim that what they call knowledge is clearly distinguishable from perceptual access to information, and that, therefore, the representation of another's knowledge is more than level-1 perspective taking. Although we remain unconvinced by this second claim, our focus rests on whether *knowledge* and *belief* are indeed categorically distinct.

The first two features put forward as essential for knowledge are that it must be (1) factive and (2) more than just true belief. In other words, you "know" something if you believe a truth for the right reasons. For instance, you can say you know that humans landed on the moon in 1969 if it is true and if you, say, watched the live transmission. If, however, you claim it was a spaceship headed for Orion that made an emergency landing on the moon, then your conviction that humans landed on the moon in 1969 would not count as knowledge, but as belief – and, well, rightly so.

We argue that the relationship between knowledge and belief is more complex and subtle. For instance, neither those who "believe" the 1969 lunar landing was accidental (because of an emergency en route to Orion) nor those who "know" it was preplanned and intentional are *knowing* these things. We might know who told us about the moon landing, but we would have no way of knowing whether their account as such is true, and the "live transmission" some of us saw could have been faked. In fact, most of us cannot even know for sure that Earth is round. Although flat-Earth proponents at least have face-evidence to back them up, we others hold our conviction to be true simply because we believe that the people who told us knew. This trust in others' knowledge is our quintessential "evidence" not only for hard-to-verify facts such as moon landings, but for almost everything we take for granted (Bender & Beller, 2019). The lion's share of our common knowledge is nothing else than belief – often plain, unverified belief – adopted from others, and hence culturally conveyed (Gatewood, 2011). Distributed knowledge and cultural transmission are key mechanisms in the process that makes human cognition unique (Bender 2020a; Tennie, Call, & Tomasello, 2009), but they come at the price of us having to trust without personal verification that what we believe we know is actually true. On the contrary, cultural consensus on agreed-upon knowledge is never complete, and diminishing consensus is one driving force for the emergence of distinct (sub)cultural truths (Gatewood, 2012).

Largely because of this cultural imprint, knowing and believing are part of a gradient rather than a simple dichotomy, contingent on the degree of uncertainty involved. Importantly, this gradient may be captured linguistically, with numerous languages even reflecting the role of cultural transmission in a much more nuanced way than English and French, the examples considered in the target article. As much as a quarter of the world's languages must qualify stated knowledge through a grammatical category called *evidentiality*. That is, speakers of these languages are obliged to specify, for every sentence they utter, the source of their information (Aikhenvald & Dixon, 2014; Chafe & Nichols, 1986), for instance whether the speaker has gained the information personally or from someone else, through direct observation, by inference, from hearsay, or assuming (Aikhenvald, 2004). Besides disqualifying English and French as sources of evidence for a universal distinction between belief and knowledge, the obligatory marking of evidentiality in many of the world's languages may also have implications for their speakers' willingness to engage in subjective activities (Luhrmann, 2011).

Junín Quechua, for instance, contains grammatical markers for indicating the source of information as being direct evidence (having seen it), conjectural, or hearsay. And, although speakers of this language make extensive use of vocabulary for talking about how things appear to be, mentalistic vocabulary is basically absent. In line with this distinct pattern of conversational topics, Junín Quechua children pass tests on representational changes and false beliefs significantly later than they pass tests on appearance-reality distinctions (Vinden, 1996; for more evidence of cross-cultural variability in the onset, unfolding, and pervasiveness of mental-state reasoning, see also Lillard, 1998; Luhrmann, 2011; Mayer & Träuble, 2013, 2015; Robbins & Rumsey, 2008; Träuble, Bender, & Konieczny, 2013).

In other words, cultural conventions and linguistic practices for defining "knowledge" affect the readiness with which people ascribe distinct mental states to other people. This generates substantial variability both in behaviors indicative of theories of mind and in teaching (Bender, 2019, 2020b) and has implications for the authors' evolutionary account, according to which "the basic capacity for knowledge representation evolved for learning from others." Although we agree that knowledge in the form of level-1 perspective taking is equally beneficial for human and nonhuman primates alike, the capacity critical for human teaching and cumulative culture would be level-2 perspective taking. Even when concerned with facts, efficiency of teaching increases with the ability to diagnose false beliefs in the learner. Unarguably, however, human teaching is even more strongly concerned with conveying beliefs, values, and practices, the high-fidelity copying of which serves to strengthen group cohesion (Legare & Nielsen, 2015). To restate a claim from the target article more precisely, "we teach others (and expect them to learn) about what we know" – and *especially* about what we believe.

In conclusion, not only is the relationship between knowledge and belief more intricate than purported in the target article (see also Polanyi, 1958), but humans have also evolved to appreciate the subtleties. Indeed, contemporary, fully enculturated humans have developed cultural as well as cognitive means to handle such subtleties with stunning finesse.

Financial support. This study was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme to the SFF *Centre for Early Sapiens Behaviour* (SapienCE), project number 262618.

Conflict of interest. None.

References

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford University Press.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Aikhenvald, A. Y., & Dixon, R. M. W. (Eds.) (2014). *The grammar of knowledge: A cross-linguistic typology*. Oxford University Press.
Bender, A. (2019). The distinct roles of theory of mind and shared intentionality for the emergence of culture. *Current Anthropology*, *60*, 182–183.
Bender, A. (2020a). The role of culture and evolution for human cognition. *Topics in Cognitive Science*, *12*, 1403–1420.
Bender, A. (2020b). What early sapience cognition can teach us: Untangling cultural influences on human cognition across time. *Frontiers in Psychology*, *11*(99), 1–6.
Bender, A., & Beller, S. (2019). The cultural fabric of human causal cognition. *Perspectives on Psychological Science*, *14*, 922–940.
Chafe, W. L., & Nichols, J. (Eds.) (1986). *Evidentiality: The linguistic coding of epistemology*. Ablex Publishing Corporation.
Gatewood, J. B. (2011). Personal knowledge and collective representations. In D. Kronenfeld, G. Bennardo, V. de Munck & M. Fischer (Eds.), *A companion to cognitive anthropology* (pp. 102–114). Blackwell.
Gatewood, J. B. (2012). Cultural models, consensus analysis, and the social organization of knowledge. *Topics in Cognitive Science*, *4*(3), 362–371.
Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences*, *19*, 688–699.
Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin*, *123*, 3–32.
Luhrmann, T. (2011). Toward an anthropological theory of mind (overview). *Suomen Antropologi: Journal of the Finnish Anthropological Society*, *36*, 5–69.
Mayer, A., & Träuble, B. E. (2013). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, *37*, 21–28.
Mayer, A., & Träuble, B. E. (2015). The weird world of cross-cultural false-belief research: A true-and false-belief study among Samoan children based on commands. *Journal of Cognition and Development*, *16*, 650–665.
Polanyi, M. (1958). *Personal knowledge*. University of Chicago Press.
Robbins, J., & Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly*, *81*, 407–420.
Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 2405–2415.
Träuble, B., Bender, A., & Konieczny, C. (2013). Human social cognition – the theory of mind research. In J. Wassmann, B. Träuble & J. Funke (Eds.), *Theory of mind in the pacific: Reasoning across cultures* (pp. 13–37). Universitätsverlag Winter.
Vinden, P. G. (1996). Junín Quechua children's understanding of mind. *Child Development*, *67*, 1707–1716.

Knowledge as commitment

Ken Binmore

Economics Department, Bristol University, Bristol BS8 1TB, UK.
k.binmore@ucl.ac.uk

doi:10.1017/S0140525X20000709, e144

Abstract

This commentary on the paper "Knowledge before belief" argues that it is not only in the cognitive sciences that knowledge should be separated into a separate category from belief, but also in rational decision theory. It outlines how knowledge-as-commitment – as distinct from knowledge-as-belief – can be built into an extension of the economic theory of revealed preference.

This note is a comment on the paper "Knowledge before belief" by Phillips et al., which argues that it is better not to regard knowledge in the cognitive sciences as a special kind of belief, but as a prior category that requires different treatment. Its aim is to endorse this conclusion, not only for the cognitive sciences, but also for economics and other disciplines that rely on rational choice theory, where the need to prioritize knowledge over belief is equally pressing. However, the approach I advocate requires departing even more than Phillips et al. from the traditional philosophical characterization of knowledge as "justified true belief" (Binmore, 2011a).

The orthodoxy in rational choice theory is Bayesianism, as formulated for applications in small worlds by Leonard Savage (1954) in his *Foundations of Statistics*. It is argued that if Alice's decisions are sufficiently consistent with each other, then she will behave as though maximizing the expected value of a utility function calculated using subjectively determined probabilities. In Savage's treatment, it is taken for granted that Alice knows the set of all possible events in advance of any analysis. In particular, when conditioning on some new information, probabilities are updated using Bayes' rule on the assumption that Alice now knows that the set of possible events is reduced to only those events consistent with her information.

What do the standard Bayesian assumptions about such information sets imply about what Alice knows? The answer is that knowledge must satisfy all the axioms that philosophers attribute to necessity (Binmore, 2009, 2011b). But a proposition is said to be necessary if and only if it is true in all possible worlds, and hence is literally impossible to refute. The foundations of Bayesianism, therefore, implicitly take for granted that knowledge is prior to belief, but users of rational choice theory nevertheless typically take for granted that knowledge should be modeled as certitude, interpreted as meaning "belief with probability one."

There is, for example, an unsettled dispute in game theory about the extent to which games more general than Chess can be solved, in principle, by working backward from all possible final positions. Is Robert Aumann (1995) right to claim that common knowledge of rationality implies backward induction in all finite games of perfect information? I think the answer is *no* if knowledge is interpreted as certitude, but trivially *yes* if knowledge is interpreted as proposed in what follows (Binmore, 2011b).

An official orthodoxy in economics is the theory of revealed (or attributed) preference (Samuelson, 1947). Suppose we are told that Alice has made certain choices. If she always chooses consistently, it is then possible to predict some of the choices she will make in the future. The Bayesian approach to belief is a late fruit of this approach. My suggestion is that it is sometimes useful to study knowledge from the same perspective, so that Alice is said to be behaving as though she knows some proposition if she always chooses as though it were true. Religious faith is a good example. I call this behavioral characterization *knowledge-as-commitment* to distinguish it from the orthodox notion of *knowledge-as-belief*.

An objection to knowledge-as-commitment is that one not only has to give up the idea that knowledge is a kind of belief, but also that Alice can only be said to know propositions that she can justify as being true. Knowledge-as-commitment captures this requirement only to the extent that attributing such knowledge to Alice requires that she never modify a proposition that she is said to know. After all, if Alice were to change her mind in ordinary life, we would say that she was mistaken to claim that she knew the proposition in the first place.

It is a hard bullet to bite that knowledge-as-commitment can survive even outright contradiction, but Donald Trump has taught us how evidence that contradicts propositions to whose

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

truth we are committed can be dismissed by calling it "fake news." However, it is not only antivaxxers and homeopaths that can be effectively modeled using the idea of knowledge-as-commitment. As Kuhn (1947) argues, normal science also explains away data that apparently refutes standard models by treating it as fake or irrelevant news. Dark matter illustrates this point in two ways. At first, physicists ignored the discovery that galaxies would not hold together if only the mass in visible stars were present. When this was no longer possible, they invented dark matter rather than meddle with the orthodox theory of gravitation.

However, my favorite example occurred when Socrates sought to explain why the Delphic Oracle had named him as the wisest man in Greece. He reasoned that it must be because he was unique in knowing that he knew nothing – by which apparently contradictory remark, I think he meant that he was committed to the view that all his beliefs were open to doubt, including those of which he was most certain. That is to say, Socrates had knowledge-as-commitment that knowledge-as-belief is always open to revision.

Conflict of interest

None.

References

Aumann, R. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior*, *8*, 6–19.

Binmore, K. (2009). *Rational decisions*. Princeton University Press.

Binmore, K. (2011a). Interpreting knowledge in the backward induction problem. *Episteme*, *8*, 248–261.

Binmore, K. (2011b). Can knowledge be justified true belief? In D. DeVidi, M. Hallet, & P. Clark (Eds.). *Logic, mathematics, philosophy: Vintage enthusiasms; essays in honor of John L. Bell* (pp. 407–412). Springer-Verlag.

Kuhn, T. (1947). *The structure of scientific revolutions*. Harvard University Press.

Samuelson, P. (1947). *The foundations of economic analysis*. University of Chicago Press.

Savage, L. (1954). *The foundations of statistics*. Wiley.

Knowledge before belief: Evidence from unconscious content

Linda A. W. Brakel[a,b]

[a]Departments of Philosophy and Psychiatry, University of Michigan, Ann Arbor, MI 48109, USA and [b]The Michigan Psychoanalytic Institute, Farmington Hills, MI 48334, USA.
brakel@umich.edu

doi:10.1017/S0140525X20000734, e145

Abstract

This commentary supports knowledge prior to belief, but from a different angle, supplementing the target article's central thesis. The target article evaluates belief-representations versus knowledge-representation in others. This commentary considers one's own unconscious knowledge, which can be prior to belief of any sort. Two examples are offered: one from clinical psychoanalysis, another involving a cognitive psychology duck/rabbit experiment.

The target article "Knowledge before belief" by Phillips et al. provides a wide array of interdisciplinary research lending a considerable empirical weight in support of the hypothesis that representations of knowledge precede representations of belief. It is my aim, with this commentary, to add to these positive findings, extending the hypothesis to a further domain – that of unconscious content – suggesting that such knowledge often proceeds belief of any sort.

To more fully explore the idea that that knowledge is prior to belief, one would have to assess whether or not the following "count" as knowledge: (a) know-how knowledge, (b) acquaintance knowledge, and (c) the capacity to discriminate among stimuli. I hold that the above three "knowledge" types, no less than clear cut representational knowledge, do count as "knowledge." This renders the question of priority to beliefs almost moot, with these three sorts of knowledge clearly preceding any manner of belief, in evolutionary, developmental, and processing terms.

However, I recognize that whether or not these three knowledge types do constitute "knowledge" is not a settled matter. Thus, and happily, the issue I raise in this commentary is far less contentious, and therefore easier to argue. Unconscious knowledge content, I aver, is often prior to any belief at all. I will offer two examples, one from clinical psychoanalytic psychotherapy, and the other from a cognitive psychology experiment. (Both are taken up in greater length in Brakel, 2010.)

Dr. X was my patient in thrice weekly psychoanalytic psychotherapy. He sat in a chair directly facing mine, some 6 feet apart. During one session, I fell asleep, albeit very briefly. For the rest of his session, I anticipated associations (direct or derivative) that would reveal not only that he saw me asleep, but that he had feelings about my lapse. But Dr. X continued as though nothing had changed … until the very next session. Immediately upon arriving, Dr. X reported the following dream: "I (Dr. X) fell asleep in our last session. You (Dr. B) did not seem to notice." He added that "there was not much feeling in the dream." I told Dr. X that his dream was clearly a reaction my having fallen asleep during our previous session; something he must've noticed. He reacted as though I were crazy, claiming with much animation: "You did not fall asleep here last time. If you had, I'd be really furious!"

Now, I was incredulous. It seemed impossible for him not to have seen me – lighting was good, line of vision clear and direct, and one's therapist is, whatever else, a salient stimulus object. The thin disguise of reversing our roles did not obscure his *unconscious knowledge of the content* – identical in waking life and in the dream – one person fell asleep right in front of the other. Furthermore, as Dr. X's knowledge of the content remained unconscious, he had absolutely no beliefs about the event he did not consciously experience. (Unconscious beliefs could be posited, but as Radford [1970, pp. 105–107] tellingly argued, this would require not only another level of inference, but more importantly the assumption that knowledge necessarily entailed belief.)

Completing the case for unconsciously knowing content proceeding any belief, consider the following: Although Dr. X did eventually believe me regarding my having fallen asleep, he never "remembered" that content. This is typical of a psychological phenomenon known as a "negative hallucination" (see Brakel, 1989a, 1989b). In negative hallucinations contents which should be supraliminal and consciously known, are clearly registered,

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

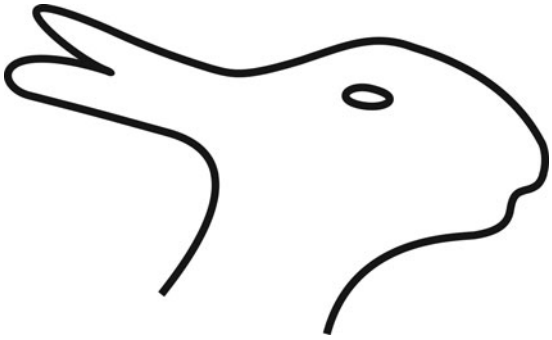


Figure 1 (Brakel). Line-drawing version of the duck/rabbit ambiguous-dual figure (Chambers and Reisberg, 1985; originally Jastrow, 1900).

but known unconsciously (non-consciously) only, much as subliminal stimuli. And, like subliminal stimuli, the contents of negative hallucinations are never recaptured consciously, and occasion no beliefs.

The second example comes from an experiment devised by Chambers and Reisberg (1985). They presented a line-drawing version of the duck/rabbit ambiguous-dual figure (Fig. 1) (originally by Jastrow, 1900) to a number of participants who were unfamiliar with it. Each participant was asked to form a mental image of the figure; the figure remaining in plain view until the participant indicated this task was complete. With each subject's mental image of the figure firmly in mind, the experimenters asked each one individually what they believed they had seen. All of the subjects who believed they'd seen both a duck and a rabbit were eliminated from the rest of the study.

Those indicating only a duck or only a rabbit had a number of further interventions, after each of which they were again to state what they believed they saw, reporting specifically on any changes in their mental/image. First, they were asked to consider other ambiguous figures. Next, they were given actual hints about the missing animal. Most of these participants stated that their image remained the same – those that believed they saw only a duck, stated that their mental image was of a duck; similarly for the only-rabbit participants. Finally, all the remaining one-animal participants were asked to draw their mental image. For the first time, as they regarded their own drawings, presto – both animals emerged. And, finally they believed the image was of a duck *and* a rabbit.

Prior to this, the only-duck participants had a belief about the duck in the image, but no belief at all about rabbits. The only-rabbit participants believed they simply saw a rabbit, but had no beliefs about ducks. However, in all of these one-animal participants, as their own drawings revealed, they undeniably had always had unconscious knowledge of the very animal about which they had no belief. After all, if you were asked to draw a duck, it would not under normal conditions, end up looking anything like a rabbit. These subjects believed only half of what they already knew.

Interestingly, this demonstrates not only that knowledge precedes belief, but also that knowledge content provides evidence for later occurring beliefs. The radical idea that knowledge provides evidence for belief, rather than the standard account that knowledge entails prior belief plus evidence, owes to Williamson (2000).

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Brakel, L. A. W. (1989a). Negative hallucinations, other irretrievable experiences, and two functions of consciousness. *The International Journal of Psychoanalysis*, *70*, 461–479.

Brakel, L. A. W. (1989b). Understanding negative hallucination: Toward a developmental classification of disturbances in reality awareness. *Journal of the American Psychoanalytic Association*, *37*, 437–463.

Brakel, L. A. W. (2010). *Unconscious knowing and other essays in psycho-philosophical analysis*. Oxford University Press.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 317–328.

Jastrow, J. (1900). *Fact and fable in psychology*. Houghton Mifflin.

Radford, C. (1970). Does unwitting knowledge entail unconscious belief? *Analysis*, *30*, 103–107.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Knowledge and the brain: Why the knowledge-centric theory of mind program needs neuroscience

Adam Michael Bricker

Department of Philosophy, Cologne Center for Contemporary Epistemology and the Kantian Tradition (CONCEPT), Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln, Germany.
adam.michael.bricker@gmail.com;
https://sites.google.com/view/adam-michael-bricker

doi:10.1017/S0140525X20001557, e146

Abstract

The knowledge-centric theory of mind research program suggested by Phillips et al. stands to gain significant value by embracing a neurocognitive approach that takes full advantage of techniques such as fMRI and EEG. This neurocognitive approach has already begun providing important insights into the mechanisms of knowledge attribution, insights which support the claim that it is more basic than belief attribution.

The knowledge-centric approach advocated by Phillips et al. represents a welcome advancement in theory of mind research, and I am in complete agreement with this proposed shift in focus. My concern, however, is that Phillips et al. have overlooked an important source of evidence available to this emerging project – the *neuroscience* of knowledge attribution. Capable of providing insights even when undetectable in behavioral measures, as well as independent lines of converging evidence, hemodynamic (e.g., functional magnetic resonance imaging [fMRI] and functional near-infrared spectroscopy [fNIRS]) and neurophysiological techniques (e.g., electroencephalography [EEG] and magnetoencephalography [MEG]) serve as powerful tools in theory of mind research. Crucially, neuroimaging has already begun to provide direct support for Phillips et al.'s central claim that knowledge attribution is more basic than belief attribution – belief attribution seems to demand neural resources that knowledge attribution does not (Bricker, 2020). All this give us compelling reason to

think that the neuroscience of knowledge attribution has a vital role to play in the nascent knowledge-centric theory of mind research program.

It is not without good reason that neuroimaging techniques have been widely employed in the effort to understand our theory of mind systems (for overviews, see Carrington & Bailey, 2009; Heleven & Van Overwalle, 2018; Mahy, Moses, & Pfeifer, 2014; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). A considerable amount of evidence indicates that the cognitive processes supporting human theory of mind capacities are both associated with identifiable neural correlates (see Heleven & Van Overwalle, 2018) and distinct from more generalized executive function in the brain (see e.g., Bradford, Brunsdon, & Ferguson, 2020; Hartwright, Apperly, & Hansen, 2015; Pacella et al., 2020; Samson, Houthuys, & Humphreys, 2015). If the knowledge-centric theory of mind program is to achieve success comparable to that of its belief-centric counterpart, this observation is key. Mental-state attributions are best understood not as cognitive, but rather as *neuro*cognitive processes.

The identifiable neural correlates of theory of mind processing enable neuroimaging techniques to provide additional lines of evidence that can converge with the findings of other methods. For example, fMRI studies indicating that the perspective taking and self-perspective inhibition components of theory of mind are largely supported by distinct regions in the brain (e.g., Hartwright et al., 2015; Özdem, Brass, Schippers, Van Der Cruyssen, & Van Overwalle, 2019; Schuwerk et al., 2014; van Der Meer, Groenewold, Nolen, Pijnenborg, & Aleman, 2011) have offered considerable support to the claim that these are indeed separate neurocognitive processes, which was initially suggested by Samson et al. primarily on the basis of lesion studies (2005).

Moreover, neuroimaging methods are especially valuable in their capacity to provide insights into theory of mind processing even when those insights aren't salient on behavioral measures. To take an example from fMRI, Hartwright et al. found differences in hemodynamic activity indicating that self-perspective inhibition during mental-state attribution is distinct from inhibition during non-mental tasks, a finding that was not detectable in their behavioral data (2015). Taking a similar example from EEG, an N400 paradigm employed by Bradford et al. revealed initial egocentric processing during the attribution of false beliefs – even when those attributions were ultimately computed successfully (i.e., altercentrically) – providing key evidence that "egocentric processing is the default perspective for information integration" in such cases (2020, p. 276). Again, this evidence was not salient in their behavioral data.

All this provides a general sense of the value of neuroimaging in theory of mind research. However, the neurocognitive findings most directly pertinent to the research program imagined by Phillips et al. come from my own EEG study (Bricker, 2020). The first step in a broader research project dedicated to understanding the neurocognitive mechanisms of knowledge attribution, the design of this study was simple, with participants varyingly judging whether a cartoon character sitting at a table knew/believed that there were two cylinders on the table. This study provided two key results: (1) There were no significant difference in response time between belief attribution and knowledge attribution. (2) Differences in P3b amplitude indicated that the belief attribution tasks demanded a level of neural resources significantly greater than that of the knowledge attribution tasks, which is most likely explained by a greater demand for self-perspective inhibition during belief versus knowledge attribution.

These findings are relevant to the account presented in the target article for at least two distinct reasons. First, these results provide additional evidence for the target article's central claim that knowledge attribution is at least as basic as belief attribution. As with the response time evidence discussed by Phillips et al. (sect. 5.1), the idea that knowledge attribution relies on something like a belief attribution stage is inconsistent with the observation of comparable response times for belief and knowledge attribution tasks. We see something similar to the neurophysiological results, which indicate that belief attribution can entail processing demands that exceed those of knowledge attribution. This again suggests that knowledge attribution is the more basic of the two processes.

However, beyond simply providing further evidence that belief attribution does not come before knowledge attribution, the results of this study also illustrate why knowledge-centric theory of mind research works best when understood as a neurocognitive endeavor, highlighting both the advantages of neurocognitive techniques outlined above. Not only did the neurophysiological results of the study provide an additional line of evidence for the conclusion suggested by behavioral measures, but these findings also offered a further insight not salient on behavioral measures – Belief attribution appears to be a more resource-intensive process than knowledge attribution, likely because of differential demands for self-perspective inhibition.

Although it is too early to speculate whether this knowledge-focused theory of mind research will ultimately attract the same attention as its belief-centric counterpart, it is clear that neurocognitive techniques have a good deal to offer this emerging project. Through the integration of behavioral and neuroimaging methods with the characterization of knowledge states offered by epistemologists (see especially sect. 2 of the target article; Bricker 2020, sect. 1.1), we stand to make significant strides toward understanding the mechanisms underlying our judgments about knowledge, which are at present still largely unknown.

Financial support. This research was supported by Sven Bernecker's Alexander von Humboldt Professor grant.

Conflict of interest. None.

References

Bradford, E., Brunsdon, V., & Ferguson, H. (2020). The neural basis of belief-attribution across the lifespan: False-belief reasoning and the N400 effect. *Cortex*, *126*, 265–280.

Bricker, A. (2020). The neural and cognitive mechanisms of knowledge attribution: An EEG study. *Cognition*, *203*, 104412–104412.

Carrington, S., & Bailey, A. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313–2335.

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, *67*, 183–192.

Heleven, E., & Van Overwalle, F. (2018). The neural basis of representing others' inner states. *Current Opinion in Psychology*, *23*, 98–103.

Mahy, C., Moses, L., & Pfeifer, J. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, *9*(C), 68–81.

Özdem, C., Brass, M., Schippers, A., Van Der Cruyssen, L., & Van Overwalle, F. (2019). The neural representation of mental beliefs held by two agents. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(6), 1433–1443.

Pacella, S., Scandola, M., Beccherle, M., Bulgarell, C., Avesani, R., Carbognin, G. ... Moro, V. (2020). Anosognosia for theory of mind deficits: A single case study and a review of the literature. *Neuropsychologia*, *148*, 107641–107641.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain*, *128*(5), 1102–1111.

Samson, D., Houthuys, S., & Humphreys, G. W. (2015). Self-perspective inhibition deficits cannot be explained by general executive control difficulties. *Cortex*, *70*, 189–201.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Schurz, R., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34.
Schuwerk, T., Döhnel, K., Sodian, B., Keck, I., Rupprecht, R., & Sommer, M. (2014). Functional activity and effective connectivity of the posterior medial prefrontal cortex during processing of incongruent mental states. *Human Brain Mapping*, *35*(7), 2950–2965.
van Der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: Neural basis of distinct processes underlying theory of mind. *NeuroImage*, *56*(4), 2364–2374.

Exchanging humpty dumpties is not a solution: Why a representational view of knowledge must be replaced with an action-based approach

Jeremy I. M. Carpendale[a] and Charlie Lewis[b]

[a]Psychology Department, Simon Fraser University, Burnaby, BC V5A 1S6, Canada and [b]Psychology Department, Lancaster University, Lancaster LA1 4YW, UK.
jcarpend@sfu.ca;
https://www.sfu.ca/psychology/about/people/profiles/jcarpend.html
c.lewis@lancaster.ac.uk;
https://www.lancaster.ac.uk/people-profiles/charlie-lewis

doi:10.1017/S0140525X20001776, e147

Abstract

In arguing for knowledge representation before belief, Phillips et al. presuppose a representational theory of knowledge, a view that has been extensively criticized. As an alternative, we propose an action-based approach to knowledge, conceptualized in terms of skill. We outline the implications of this approach for children's developing social understanding, beginning with sensorimotor interaction and extending to the verbal level.

In their target article, Phillips and colleagues focus on the important problem of how people understand others' minds, and they argue that children's ability to represent knowledge develops before their ability to represent beliefs. They recognize that in focusing on knowledge they are obliged to be clear about their assumptions. And in their "call to arms," they state that their goal is to reach a "deeper understanding of the nature of knowledge." We applaud this goal because if we are concerned with how children learn about the world, including other people, then we must be aware of our assumptions regarding knowledge (Chapman, 1999). Yet, in Phillips et al.'s following sentence, and indeed throughout, it is clear that they fail to explicate or assess their own assumptions and simply assume a representational view, according to which knowledge is based on representations that match the world. This view is also labeled a correspondence or copy theory (Piaget, 1970), or spectator notion of knowledge (Dewey, 1960), and has been extensively critiqued (e.g., Bickhard, 2009; Carpendale & Lewis, 2004, 2006, 2021; Dewey, 1960; Piaget, 1970). Russell (1992) refers to representation as a humpty dumpty term because it can be used in any way desired to mean absolutely nothing.

This is a passive view of knowledge which, it has often been argued, presupposes "what it is meant to explain" (Chapman, 1999, p. 31). That is, from this perspective knowledge consists of representing or forming a copy of the world. But, in that case, the only way to check our knowledge is by forming another copy to compare it to because we do not have direct access to the world. This does not solve the problem, and thus we cannot tell if the copy is accurate. Yet becoming aware of errors and learning from them is essential in understanding development (Bickhard, 2009). Thus, this view already presupposes knowledge when that is what it is meant to explain, and thus the account is inadequate. Instead of taking for granted the ability to re-present and think about what is not immediately there, we view this as an important challenge to explain. We follow others in not taking our biggest problem and making it our premise (Bickhard, 2009; Piaget, 1970).

Understanding how children learn about knowledge and beliefs depends on an adequate conception of the more general issue of how children develop knowledge. As an option that deals with the criticism of the representational view, we endorse an action-based or constructivist approach, according to which knowledge is conceptualized in active terms of learning what can be done with the world, that is, of its interactive potential. Thus, knowledge is viewed as skill rather than "representation" (Bibok, Carpendale, & Lewis, 2008). Instead of assuming that the child possesses a pre-existing mind and faces "the problem of other minds" when trying to understand others, human infants develop within a social system. They gain mastery of the world through coming to anticipate what happens when they do things. For example, they learn how to reach for and grasp objects of interest. Because this action typically occurs in a social context, infants' interests are manifest for other people. Others may respond if the infant is not initially successful in reaching the desired object. Thus, their goal may be reached not because they attain it but because it prompts facilitation from a caregiver. Infants can then gradually become aware that their actions are of significance for others, that is, they learn that they are communicating (Mead, 1934). Infants can then begin to communicate intentionally, first with gestures and later with the addition of words (Carpendale, 2018).

From this perspective, infants first form of knowledge of other people in terms of an anticipation of their action based on their prior activity. This is achieved first at the practical sensorimotor level. It is this form of knowledge that can be evident in infant false-belief tasks (Carpendale & Lewis, 2015). This means of understanding others is also evident in research with chimpanzees, which shows that in competitive situations they can anticipate dominant co-specifics' actions such as moving toward food depending on whether they have seen it (e.g., Tomasello, Call, & Hare, 2003).

We believe that the distinction between knowledge and belief that Phillips et al. discuss refers to much later distinctions involving language. Children come to use terms referring to others' anticipated action in terms of their knowledge or beliefs, as in "she thought that…" or "she knows that…" To be able to talk about and think about others in terms of knowledge and belief depends on learning how to use such words. These words refer to patterns of human activity not to inner mental entities that mental-state words are mapped onto (Canfield, 2007; Wittgenstein, 2009). Children gradually learn some of the many ways in which words such as think and know are used. To consider just a few early uses of such terms, the word "know" can be used in various ways – in relation to whether someone has seen something, to refer to someone's ability or lack of ability, to deny responsibility as in "I didn't know," or to avoid answering

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

as in "I don't know." The word "think" can be used to modulate certainty or to refer to a process of decision-making.

We have focused on a more fundamental problem with Phillips et al.'s position than the concern with whether children "compute" belief representations before or after knowledge representations. This is because their way of conceptualizing both belief and knowledge is problematic because of being based on a problematic conception of knowledge and mind.

What we are suggesting as an alternative to Phillips et al. is not empiricism, which neglects the activity of the subject (Piaget, 1970, 1972), but instead is an action-based, process-relational approach (Carpendale, Hammond, & Atwood, 2013b). Our focus on activity should not be mistaken for behaviorism. Instead, we begin from action and interaction in order to explain psychological development (Carpendale, Atwood, & Kettner, 2013a; Carpendale & Lewis, 2015).

Conflict of interest

None.

References

Bibok, M. B., Carpendale, J. I. M., & Lewis, C. (2008). Social knowledge as social skill: An action based view of social understanding. In U. Müller, J. I. M. Carpendale, N. Budwig & B. Sokol (Eds.), *Social life and social knowledge: Toward a process account of development* (pp. 145–169). Taylor Francis.

Bickhard, M. (2009). The interactivist model. *Synthese*, *166*, 547–591. doi:10.1007/s11229-008-9375-x.

Canfield, J. V. (2007). *Becoming human: The development of language, self, and self-consciousness*. Palgrave Macmillan.

Carpendale, J. I. M. (2018). Communication as the coordination of activity: The implications of philosophical preconceptions for theories of the development of communication. In A. S. Dick & U. Müller (Eds.), *Advancing developmental science: Philosophy, theory, and method* (pp. 145–156). Routledge, Taylor & Francis.

Carpendale, J. I. M., Atwood, S., & Kettner, V., (2013a). Meaning and mind from the perspective of dualist versus relational worldviews: Implications for the development of pointing gestures. *Human Development*, *56*, 381–400.

Carpendale, J. I. M., Hammond, S. I., & Atwood, S. (2013b). A relational developmental systems approach to moral development. In R. M. Lerner & J. B. Benson (Eds.), *Embodiment and epigenesis: Theoretical and methodological issues in understanding the role of biology within the relational developmental system* (Vol. *1*, pp. 105–133). *Advances in child development and behavior, vol. 45*. Academic Press.

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*, 79–96.

Carpendale, J. I. M., & Lewis, C. (2006). *How children develop social understanding*. Blackwell Publishers.

Carpendale, J. I. M., & Lewis, C. (2015). The development of social understanding. In L. Liben & U. Müller (Eds.), Vol. 2: Cognitive processes, R. Lerner (editor-in-chief), *7th Edition of the Handbook of child psychology and developmental science* (pp. 381–424). Wiley-Blackwell.

Carpendale, J. I. M., & Lewis, C. (2021). *What makes us human? How minds are constructed in relationships*. Routledge.

Chapman, M. (1999). Constructivism and the problem of reality. *Journal of Applied Developmental Psychology*, *20*(1), 31–43.

Dewey, J. (1960). *On experience, nature, and freedom*. The Bobbs-Merrill Company, Inc.

Mead, G. H. (1934). *Mind, self and society*. University of Chicago Press.

Piaget, J. (1970). *Genetic epistemology*. Norton.

Piaget, J. (1972). *Psychology and epistemology: Towards a theory of knowledge*. Viking Compass Book.

Russell, J. (1992). The theory theory: So good they named it twice? *Cognitive Development*, *7*, 485–519.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states: The question is which ones and to what extent. *Trends in Cognitive Science*, *7*, 153–156.

Wittgenstein, L. (2009). *Philosophical investigations*. (Revised 4th ed., translation by G.E.M. Anscombe, P.M.S. Hacker & J. Schulte). Wiley-Blackwell. (original work published 1953).

Relational mentalizing after any representation

Eliane Deschrijver[a,b]

[a]Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000, Ghent, Belgium and [b]School of Psychology, University of New South Wales (UNSW), Library Walk, Kensington, NSW 2033, Australia.
e.deschrijver@unsw.edu.au; www.elianedeschrijver.com

doi:10.1017/S0140525X20001417, e148

Abstract

Autistic, developmental, and nonhuman primate populations fail tasks that are thought to involve attributing beliefs, but not those thought to reflect the representation of knowledge. Instead of knowledge representations being more basic than belief representations, *relational mentalizing* may explain these observations: The tasks referred to as reflecting "belief" representation, but not the "knowledge" representation tasks, are social conflict designs. They involve mental conflict monitoring *after* another's mental state is represented – with effects that need to be accounted for.

The ability to represent others' mental states has been thought of as social cognition's primary building block for over 40 years (Apperly, 2010; Premack & Woodruff, 1978). Phillips et al. take this one step further in their impressive study of interdisciplinary thought: They argue that knowledge representation is a fundamentally different and more basic process than belief representation. This claim is based, in large part, on the observation that nonhuman primates (sect. 4.1 in Phillips et al.) as well as developmental (sect. 4.2) and clinical (sect. 4.4) populations show atypical behavior in false-belief tasks, despite their intact performance in tasks that may tap into the representation of knowledge.

A basic premise in order to accept the authors' argument is that atypical performance in a false-belief task arises from a lack of ability to represent another's belief (Dennett, 1978). But is this necessarily the case? We recently pointed out in our newly developed relational mentalizing framework (Deschrijver & Palmer, 2020) that although passing a false-belief task is sufficient to conclude a belief representation ability to be *present*, failing one does not yield evidence for it to be *absent*. False-belief tasks are social conflict designs: The other typically holds a mental state that is manipulated to be irreconcilable with their own (e.g., they may think that the ball is in the basket, whereas you think it is in the box). In order to resolve the mental conflict that arises from having to represent both mental states, a neural mechanism may need to inhibit one of the competing representations before focusing on the other (i.e., mental conflict monitoring; Deschrijver & Palmer, 2020). However, for example in a young child, if the mechanism fails to inhibit their own misaligned mental state *despite being able to represent both states*, it may manifest as an inability to verbalize or show sensitivity to the other's false belief. Even neurotypical adults, who undoubtedly represent others' mental states, can make errors in a false-belief task if they fail to suppress their own mental representation, suggesting that the mechanism indeed exists (Keysar, Lin, & Barr, 2003; for other evidence, see Deschrijver & Palmer, 2020). Ineffective mental conflict monitoring may also be reflected

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

in *more* interference by the other's belief if the task requires you to inhibit the representation of the other's mental state to focus on your own, as reported in some adults on the spectrum (Deschrijver, Bardi, Wiersema, & Brass, 2016). Developmental (e.g., Kovács et al., 2010), nonhuman primate (e.g., Martin & Santos, 2014), and autistic (e.g., Deschrijver et al., 2016) populations may thus show atypical results in false-belief tasks even while representing the other's belief.

The tasks identified by Phillips et al. as showing intact "knowledge" representation in young, autistic, and nonhuman primate populations broadly follow two methodological designs: First, they assess the understanding of the another's *ignorance* (e.g., Flombaum & Santos, 2005; Luo & Johnson, 2009; Santos, Nissen, & Ferrugia, 2006), meaning that the other does not have knowledge about the object of interest. Second, they investigate whether one understands that the other has knowledge (a mental state that is true), about the object's location with the observer either having or not having access to this other's knowledge (e.g., Behne, Liszkowski, Carpenter, & Tomasello, 2012; Luo & Johnson, 2009). To solve these tasks successfully, there is no need for the brain to engage in mental conflict monitoring regarding the object's location: There is no other-related mental state to be represented (and, therefore, it cannot clash with one's own), or the other's mental state *aligns* with the own. Populations that do represent others' mental states, but are unable to deal with mental conflict, should hence not experience any difficulties. Consistent with this, difficulties arise if the other's (unknown) knowledge starts misaligning with the own understanding of the world (e.g., Krachun, Carpenter, Call, & Tomasello, 2009), and when the design temporarily involves a manipulation of conflict between the other's and the participant's understanding of reality (e.g., where the object is ostentatiously relocated without the other seeing it, but then put back; Fabricius, Boyer, Weimer, & Carroll, 2010; Gettier, 1963; Horschler, Santos, & MacLean, 2019; Kaminski, Call, & Tomasello, 2008). What Phillips et al. consider to be a true belief design with representations that do not doesn't qualify "knowledge," thus involves a short-term manipulation of mental conflict as well. This means that atypical results in such "true belief" tasks may be attributable to mental conflict monitoring rather than belief representation issues in these populations, too. To show that the distinction between representing another's belief versus knowledge consists of more than semantics, the field may thus need to show that populations such as young children, nonhuman primates, and individuals on the spectrum continue to perform well in "knowledge" tasks that *do* contain mental conflict (e.g., Samson et al., 2010; see Table 3 in Deschrijver & Palmer, 2020, for the most optimal dependent measures), and fail in (true and false) "belief" tasks that don't. Mental conflict monitoring is also likely effortful, using cognitive resources and showing relationships with executive functions (Carlson, 2010; Carlson, Mandell, & Williams, 2004a; Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Moses, & Claxton, 2004b). This may result in seemingly less automatic or slower responses in false-belief designs versus those knowledge designs that do not involve mental conflict (see sects 4.3 and 5.3 in Phillips et al.). From all these arguments, it follows that a differential performance of autistic, developmental, and non-human primate populations in "belief" versus "knowledge" tasks could be seen not as a consequence of these tasks tapping into two different types of representations that are differentially affected (i.e., "knowledge" versus "beliefs"), but rather as an indication that these populations have issues with solving mental conflict instead of with attributing to others any representation *per se*.

Representing another's belief versus knowledge is thus not as easily dissociable as the authors may want to portray: If the two representation mechanisms are truly distinct, shouldn't one always be able to assess the truthfulness of another's mental state *before* representing it? This seems challenging in the real world, where others' mental states are more complex than the ones typically used in the mentalizing domain. Regardless of which party holds the facts, however, mental conflict may arise *after* representing another's position if it's misaligned. In sum, when taking into account mental conflict monitoring, the idea that the representation of beliefs and knowledge are two fundamentally different things, with one being more basic than the other, may be on shaky grounds.

Financial support. The author received funding from the Research Foundation Flanders – FWO (postdoctoral fellowship).

Conflict of interest. None.

References

Apperly, I. A. (2010). *Mindreaders: The cognitive basis of "theory of mind."* Psychology Press. http://doi.org/10.4324/9780203833926.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, *30*, 359–375.

Carlson, S. M. (2010). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*, 37–41. http://dx.doi.org/10.1207/s15326942dn2802_3.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004a). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*, 1105–1122. http://dx.doi.org/10.1037/0012-1649.40.6.1105.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032–1053. http://dx.doi.org/10.1111/1467-8624.00333.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*, 73–92. http://dx.doi.org/10.1002/icd.298.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004b). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*, 299–319. http://dx.doi.org/10.1016/j.jecp.2004.01.002.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*, 568–570. http://dx.doi.org/10.1017/S0140525X00076664.

Deschrijver, E., Bardi, L., Wiersema, J. R., & Brass, M. (2016). Behavioral measures of implicit theory of mind in adults with high functioning autism. *Cognitive Neuroscience*, *7*, 192–202. http://dx.doi.org/10.1080/17588928.2015.1085375.

Deschrijver, E., & Palmer, C. (2020). Reframing social cognition: Relational versus representational mentalizing. *Psychological Bulletin*, *146*(11), 941–969. https://doi.org/10.1037/bul0000302.

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.

Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, *15*, 447–452.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

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*, 72–80.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*, 224–234.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41. http://dx.doi.org/10.1016/S0010-0277(03)00064-7.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibly to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149.

Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition*, *130*, 300–308. http://dx.doi.org/10.1016/j.cognition.2013.11.016.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526. http://dx.doi.org/10.1017/S0140525X00076512.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other

people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–1266. doi: 10.1037/a0018729.
Santos, L. R., Nissen, A. G., & Ferrugia, J. (2006). Rhesus monkeys (*Macaca mulatta*) know what others can and cannot hear. *Animal Behaviour*, *71*, 1175–1181.

Do "knowledge attributions" involve metarepresentation just like belief attributions do?

Rachel Dudley and Ágnes Melinda Kovács

Department of Cognitive Science, Cognitive Development Center, Central European University, Budapest, Oktober 6 u. 7, 1051, Hungary.
dudleyr@ceu.edu; kovacsag@ceu.edu
https://sites.google.com/site/rachelelainedudley;
https://people.ceu.edu/agnes-melinda_kovacs

doi:10.1017/S0140525X20001594, e149

Abstract

The authors distinguish knowledge and belief attributions, emphasizing the role of the former in mental-state attribution. This does not, however, warrant diminishing interest in the latter. Knowledge attributions may not entail mental-state attributions or metarepresentations. Even if they do, the proposed features are insufficient to distinguish them from belief attributions, demanding that we first understand each underlying representation.

The authors argue for a distinction between so-called "knowledge attributions" and belief attributions, where the former are defined by four unique features: (1) necessarily true content, (2) difference from true belief, (3) compatibility with egocentric ignorance, and (4) modality-generality. The authors suggest that research on theory of mind should shift toward investigating "knowledge attributions" because they are more basic. We agree that there may be two different kinds of representations underlying the authors' distinction, and that one kind of representation may be more basic (even if it may not be developmentally prior). However, we do not agree with the unique features which they take to characterize "knowledge attributions," or with the claim that these are core representations for mental-state attribution.

The authors suggest that *belief* plays a central role in the field since Premack and Woodruff (1978) and associated commentaries, and position themselves in opposition to this focus on belief. Instead, we argue that *metarepresentation* is at the core of attributing representational mental states; focus fell on false belief because it provides a stringent diagnostic for metarepresentation. In our view, most cases of "knowledge attributions," as defined by the authors, could be explained without appealing to metarepresentation or even to attribution in the traditional sense. Although the authors provide no details on what they take the underlying format to be, it may merely involve two simple representations: our own reality representation and a copy of it to be used for others (Phillips & Norby, 2019). This kind of non-metarepresentational format would allow one to encode the most up-to-date state of reality for oneself and for others, thereby enabling predictions about who will act in a reality-congruent way or be a reliable source of information. In case there are such representations, more research is needed to characterize their format and their role in cognition. Nevertheless, they should not become the main focus in theory of mind as they do not seem to explain a wide range of phenomena, which rely on tracking how other individuals represent the world without necessarily corresponding to reality.

Other cases of "knowledge attributions," as defined by the authors, could truly be understood as mental-state attributions involving metarepresentation but we would argue that they are indistinguishable from belief attributions. In particular, we find the four features used by the authors to be insufficiently predictive. Specifically, one subset of these features is compatible with belief attributions (3 and 4); and the other subset seems ad-hoc, especially from the perspective of cognitive mechanisms (1 and 2). On the first point, studies have shown that false-belief attribution also allows for egocentric ignorance (3) in many populations (Biervoye, Meert, Apperly, & Samson, 2018; Call & Tomasello, 1999; Krachun, Carpenter, Call, & Tomasello, 2009; Samson, Apperly, Kathirgamanathan, & Humphreys, 2005). In fact, attribution under egocentric ignorance may be one of the best illustrations of the metarepresentational format because attributed beliefs and their content can be manipulated independently (Leslie, 1987). Similarly, false-belief attribution also exhibits modality-generality (4) because it allows for integration of content from different perceptual sources (Song, Onishi, Baillargeon, & Fisher, 2008; Tauzin & Gergely, 2018). On the second point, where the features are indeed incompatible with belief attribution, they merely stipulate the difference from true beliefs (2), or they seem to depend on facts in the external world to secure true contents (1), which may be better captured by epistemologists than psychologists (Ichikawa & Jenkins, 2017).

As a general point, we see no way to distinguish these "knowledge attributions" from certain belief attributions on the basis of their truth, at least as these attributions are currently understood within cognitive science. It may be more fruitful to recast the distinction within perception instead of truth, for conceptual as well as empirical reasons. From a conceptual perspective, we cannot imagine how a non-verbal creature could attribute knowledge versus belief in a real-world scenario *without* appealing to physical cues to perceptual access as a proxy for knowledge. And from an empirical perspective, when working with non-verbal participants such as infants and nonhuman primates, we are limited to testing contexts where beliefs/knowledge are formed on the basis of perceptual access. Despite this, data suggest that uninterrupted perceptual access is the only relevant factor for nonhuman primates, although this does not seem to be the case for children: Independent of their success in false-belief conditions, children's performance is modulated by multiple factors (Kaminski, Call, & Tomasello, 2008). Future research should try to understand the nuanced factors that contribute to human mental-state attribution as opposed to collapsing it into the two-way distinction that may better explain primate findings.

Similar to the authors, we will end by addressing the role of mental-state attribution in learning. Unlike the authors, we argue that it is actually belief attributions which undergird learning for both conceptual and empirical reasons (Kampis, Somogyi, Itakura, & Király, 2013). Given that we cannot see how a young child could hope to distinguish knowledge from mere true belief unless they use uninterrupted perceptual access as a proxy, belief attributions should be equally good to motivate learning. Furthermore, experimental research supports that children and adults rely on *false* beliefs to learn about the physical world when they themselves lack perceptual access (to locate a target; Biervoye et al., 2018; Call and Tomasello, 1999; Krachun et al., 2009; Samson et al., 2005); and even infants

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

use *false*-belief attributions to acquire person-specific ("She prefers object A"; Luo, 2011) or *generalizable* information ("Object A is preferable"; Kampis et al., 2013).

Differentiating representations of reality from mental-state attributions is important to furthering our understanding of the mind. We need to better understand each kind of representation, and its role in broader cognition. But this should not undermine or replace research into theory of mind, which should remain centered around metarepresentation and mental-state attribution. Within this domain, we need better theoretical constructs to distinguish flavors of mental-state attribution, if and when we wish to distinguish them at all. Even after decades, we barely understand the processes and representations involved in this central aspect of human cognition.

Financial support. This research is supported by a McDonnell Foundation Network Grant: "The Ontogenetic Origins of Abstract Combinatorial Thought."

Conflict of interest. None.

References

Biervoye, A., Meert, G., Apperly, I. A., & Samson, D. (2018). Assessing the integrity of the cognitive processes involved in belief reasoning by means of two nonverbal tasks: Rationale, normative data collection and illustration with brain-damaged patients. *PLOS ONE*, *13*(1), e0190295.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, *70*(2), 381–395.

Ichikawa, J., & Jenkins, C. (2017). On putting knowledge "first." In J. A. Carter, E. C. Gordon, & B. Jarvis (Eds.), *Knowledge first: Approaches in epistemology and mind* (pp. 113–131). Oxford University Press.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234.

Kampis, D., Somogyi, E., Itakura, S., & Király, I. (2013). Do infants bind mental states to agents? *Cognition*, *129*(2), 232–240.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535.

Leslie, A. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*(4), 412–426.

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition*, *121*(3), 289–298.

Phillips, J., & Norby, A. (2019). Factive theory of mind. *Mind & Language*, *36*(1), 1–24.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain*, *128*(5), 1102–1111.

Song, H. J., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, *109*(3), 295–315.

Tauzin, T., & Gergely, G. (2018). Communicative mind-reading in preverbal infants. *Scientific reports*, *8*(1), 1–9.

Representing knowledge, belief, and everything in between: Representational complexity in humans and other apes

Kresimir Durdevic[a] and Christopher Krupenye[b,c]

[a]School of Psychology & Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK; [b]Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD 21218 and [c]Department of Psychology, Durham University, Durham DH1 3LE, UK.
kd96@st-andrews.ac.uk; krupenye@jhu.edu
http://christopherkrupenye.weebly.com/

doi:10.1017/S0140525X20001855, e150

Abstract

Building on Phillips and colleagues' case for the primacy of knowledge, we advocate for attention to diversity in mentalizing constructs within, as well as between, knowledge and belief. Ultimately, as great apes and other animals show, the development and evolution of theory of mind may reflect a much greater range of incremental elaborations of representational or computational complexity.

The target article (Phillips et al.) presents impressively diverse evidence for carving species-wide mental attitude attribution into two principal representational capacities: ascription of factive knowledge and non-factive beliefs. Although we find evidence for the primacy of knowledge compelling, we believe there is room to carve the theoretical space further into a diverse representational spectrum that characterizes the development and evolution of factive and non-factive theory of mind. Our strategy, similar to others' (Nagel, 2017; Phillips & Norby, 2021), is to start from the attributor's egocentric representation of reality and take minimal steps to build toward more charitable inclusion of the altercentric agent. We then address the case of great ape mindreading.

Knowledge attribution may pose the least extra demand on an attributor's egocentric representation. The attributor can get away with simply extending her personal representation to an altercentric agent. Presumably, noting the agent's presence and her epistemic contact (perceptual or inferential) with something in the attributor's egocentric reality is enough. From this, the attributor can generate a positive attribution of mental state with some content and, hence, a degree of separation between the ego and altercentric. The behavioral expectation for the agent will be based on this generalization of the egocentric reality.

Altercentric ignorance describes a situation in which the agent fails to witness something in the attributor's egocentric representation. The attributor has at least two ways of handling this situation. Agents' lack of epistemic contact with some event (e.g., "she did not see or could not know x") may amount to no attribution and no behavioral expectation. Alternatively, the attributor can omit from her representation what the agent "missed" and attribute that modified representation (Phillips & Norby, 2021). Phillips and Norby (2021) qualify ignorance as knowledge-like, with both requiring the attributor to hold the personal representation and an altercentric copy in parallel. Martin and Santos (2016) qualified this separation of the ego and altercentric as already akin to false beliefs, and suggested, in contrast, that the most basic response to ignorance is a simple *failure* to track the altercentric. Both characterizations of ignorance constitute forms of mindreading with different levels of representing the altercentric.

Egocentric ignorance constitutes cases in which the agent has privileged epistemic contact in comparison with the attributor – for example, a student, nervous about her exam results in an email, asks her friend to read it on her behalf. The student knows *that* her friend knows the results, but not *what* results. In this case, the attributor can no longer rely on her egocentric reality to handle the agent's. Hence, the egocentric ignorance situation requires a new way to represent the agent's altercentricity – the agents no longer succeeds or fails to know what the attributor knows but rather they have epistemic opportunities of their own. In egocentric ignorance cases, the content of the agent's attitude may arise independent of the attributor's reality but, critically, it is not specified (because the attributor does not know it). It is,

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

therefore, not quite belief-like in the sense that there is no contrast in content of the egocentric and altercentric representation that could be interpreted as counterfactual (sensu Phillips & Norby, 2021). Egocentric ignorance, thus, constitutes an intermediate representational sophistication – more than knowledge, shy of belief.

False-belief representation is the first situation in which the altercentric representation must not only be held as independent, but the content of this altercentric state must also be updated dynamically. The attributor first tracks the agent as knowledgeable (that "object is in box A"; e.g., as they co-witness a hiding event), then ignorant (that "object is not in box A, object is in box B"; e.g., as the object is displaced in the agent's absence).

True belief representation, at least as exemplified in the Gettier case (Gettier, 1963), constitutes an even more demanding computational task. Here, agents co-witness an object's initial hiding and then, in the agent's absence, the object is temporarily removed and then returned to the same location (Horschler, Santos, & MacLean, 2019; Kaminski, Call, & Tomasello, 2008). Consequently, the requisite updating process involves reconciliation of two contrasting altercentric mental states (knowledge and ignorance) with identical content (object is in location A).

We can now sketch a "scale" of cognitive complexity across the factive and non-factive mental attribution spectrum. Provided she can individuate agents, a hypothetical mindreading attributor can, at the most basic level, track epistemic contact or lack thereof. Egocentric reality alone is sufficient for tracking this minimal notion of knowledge and ignorance. Next, she can attribute copies of her egocentric information, and, in the case of altercentric ignorance, modify them. In egocentric ignorance contexts, the altercentric agent is granted a new epistemic status without the cognitive cost of attributing fully specified content to her. For belief tracking, the content of this altercentric representation is specified and updated. This process is even more computationally challenging in paradigmatic cases of true belief attribution. Finally, aspectuality may (or may not) impose further representational complexity (e.g., Perner, Huemer, & Leahy, 2015; but see Rakoczy, Bergfeld, Schwarz, & Fizke, 2015).

If the present hierarchical characterization is correct, the predictions are clear. Children or animals proficient in more sophisticated abilities should master simpler ones (e.g., Fig. 1 of Krupenye, 2020). Great apes, for example, have shown competence on several recent false-belief tasks (Buttelmann, Buttelmann, Carpenter, Call, & Tomasello, 2017; Kano, Krupenye, Hirata, Tomonaga, & Call, 2019; Krupenye, Kano, Hirata, Call, & Tomasello, 2016), raising the possibility that they may indeed track beliefs. Current evidence is tentatively consistent with the proposed complexity scale, suggesting that apes also track knowledge and altercentric ignorance (Hare, Call, & Tomasello, 2006; Karg, Schmelz, Call, & Tomasello, 2015), and potentially egocentric ignorance (Call & Tomasello, 1999; Krachun, Carpenter, Call, & Tomasello, 2009), but perhaps not Gettier's true beliefs (Kaminski et al., 2008). Cases of egocentric ignorance, in particular however, deserve further, more targeted tests. Broader efforts in humans and nonhumans also demand new tasks that carefully tease apart attribution of knowledge and true belief, and of knowledge, egocentric ignorance, and false belief. Together, these developments will clarify the family, or hierarchy, of factive and non-factive theory of mind.

Financial support. KD was supported by UK Economic and Social Research Council studentship 2267016 and a St Leonards Research Scholarship from the University of St Andrews, and CK by European Commission Marie Sklodowska-Curie fellowship MENTALIZINGORIGINS.

Conflict of interest. None.

References

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). Great apes distinguish true from false beliefs in an interactive helping task. *PLOS ONE*, *12*(4), e0173793. https://doi.org/10.1371/journal.pone.0173793.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, *70*(2), 381–395. https://doi.org/10.1111/1467-8624.00028.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123. JSTOR. https://doi.org/10.2307/3326922.

Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition*, *101*(3), 495–514. https://doi.org/10.1016/j.cognition.2005.01.011.

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*, 72–80. https://doi.org/10.1016/j.cognition.2019.04.012.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234. https://doi.org/10.1016/j.cognition.2008.08.010.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences*, *116*(42), 20904–20909. https://doi.org/10.1073/pnas.1910095116.

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015). Chimpanzees strategically manipulate what others can see. *Animal Cognition*, *18*(5), 1069–1076. https://doi.org/10.1007/s10071-015-0875-z.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x.

Krupenye, C. (2020). The evolution of mentalizing in humans and other primates. In M. Gilead & K. Ochsner (Eds.), *The neural basis of mentalizing: A social-cognitive and affective neuroscience perspective* (pp. 107–129). Springer Press. doi: 10.1007/978-3-030-51890-5.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110.

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375–382. https://doi.org/10.1016/j.tics.2016.03.005.

Nagel, J. (2017). Factive and nonfactive mental state attribution. *Mind & Language*, *32*(5), 525–544. https://doi.org/10.1111/mila.12157.

Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition*, *145*, 77–88. https://doi.org/10.1016/j.cognition.2015.08.006.

Phillips, J., & Norby, A. (2021). Factive theory of mind. *Mind & Language*, *36*, 3–26. https://doi.org/10.1111/mila.12267.

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.

The role of epistemic emotions in learning from others

Asli Erdemli, Catherine Audrin, and David Sander

Swiss Center for Affective Sciences, University of Geneva, Campus Biotech, Chemin des Mines 9, 1202 Geneva, Switzerland.
asli.erdemli@unige.ch; catherine.audrin@unige.ch; david.sander@unige.ch
https://www.unige.ch/fapse/e3lab/members1/phd-candidates/asli-erdemli/
https://www.unige.ch/fapse/e3lab/members1/post-docs/dr-catherine-audrin/
https://www.unige.ch/fapse/e3lab/director/

doi:10.1017/S0140525X20001624, e151

Abstract

Phillips et al. discuss whether knowledge or beliefs are more basic representations of others' minds, focusing on the primary function of knowledge representation: learning from others. We discuss links between emotion and "knowledge versus belief," and particularly the role of emotions in learning from others in mechanisms such as "social epistemic emotions" and "affective social learning."

Current emotion research emphasizes the existence of a specific family of emotions whose objects are knowledge or the process of knowledge generation/acquisition: these emotions are called "epistemic emotions" or "knowledge emotions" (Brun et al., 2008; Morton, 2010). Unlike achievement emotions, they are not related to the success or failure at a certain task but to the epistemic content or process itself (Pekrun & Stephens, 2012). Examples of currently studied epistemic (or "knowledge") emotions are surprise, curiosity, enjoyment, confusion, anxiety, frustration, and boredom (see Pekrun, Vogl, Muis, & Sinatra, 2017, for the Epistemically-Related Emotions Scale). Epistemic curiosity (i.e., epistemic interest), probably the most widely studied epistemic emotion yet, activates reward-related regions in the brain (Gruber, Gelman, & Ranganath, 2014; Kang et al., 2009). It enhances memory for the content the individual was curious about (Kang et al., 2009) but also of incidental information presented during high states of epistemic curiosity (Gruber et al., 2014). Epistemic curiosity creates additional knowledge exploration and better knowledge acquisition (Ainley, 2017; Wade & Kidd, 2019). Some studies focused on the antecedents of curiosity (Connelly, 2011; Silvia, 2005) to find out what creates curiosity for knowledge in humans. Others have even shown that healthy adults would risk electrical shocks to learn about curiosity-inducing knowledge (Lau, Ozono, Kuratomi, Komiya, & Murayama, 2020).

Although epistemic emotions are personal affective experiences elicited by knowledge, we are not aware of any study that specifically focused on how learners feel epistemic emotions *about knowledge they attribute to others*. For achievement emotions, there is a category of emotions called "social achievement emotions" (e.g., admiration, envy, contempt, and empathy), which is about the success and failure of others (see Pekrun & Linnenbrink-Garcia, 2014). By analogy, we propose the existence of "social epistemic emotions" which refer to the epistemic emotions whose objects are knowledge represented in others. Such emotions, although felt at the first person, would be about third-person knowledge, and be a driving force supporting what Philips et al. consider as the primary function of knowledge representation, namely learning from others. A non-exhaustive list of social epistemic emotions could be: surprise (e.g., the learner is surprised by the representation of the knowledge attributed to the other), curiosity (e.g., the learner feels intrinsically motivated to learn more about the represented knowledge), confusion (e.g., the learner attributes to the social source a knowledge representation contrary to their own prior knowledge, and is experiencing cognitive conflict as a result), and admiration (e.g., the learner is impressed by the quality and/or quantity of knowledge they represent the social source to have). The study of social epistemic emotions should include a broad variety of social sources (e.g., teachers, caregivers, and peers) that play a considerable role in knowledge acquisition (see Harris, Bartz, & Rowe, 2017, for a review on how children turn to their social environment to learn about the world). Social epistemic emotions should help the learner select relevant social sources of knowledge (e.g., through trust-related mechanisms), energize behaviors of knowledge-seeking (e.g., through social interactions) which would eventually lead to actual learning from others. For instance, teacher competence enhances student interest and achievement (Fauth et al., 2019). Research could investigate whether this effect is mediated by the student's representation of the teacher's knowledge. Examples of frameworks in which social epistemic emotions could play a particularly important role are peer-to-peer learning, tutor-student learning, group assignments, debates, and so on.

In contrast to the growing literature concerning the nature and functions of epistemic emotions and the role these emotions play in knowledge acquisition, to the best of our knowledge, there is no category of emotion suggested to have belief – rather than knowledge – as their objects. In particular, we are not aware of any study that aimed at comparing emotions elicited by "knowledge in others" to emotions elicited by "belief in others." We speculate that, because beliefs can be false, if a learner comes across a social source who explicitly expresses their representation as a belief (e.g., by saying "I believe that p"), they will feel less curious and motivated to explore further that representation than if they express it as a knowledge (e.g., " I know that p").

In addition to what has been said on the role of knowledge and belief representations in learning, links may be considered with respect to the robust and growing body of literature on learning from the emotions of others. Affective social learning (Clément & Dukes, 2017), of which "social appraisal" (Fischer, 2019; Manstead & Fischer, 2001, 2017) and "social referencing" (Klinnert, Campos, Sorce, Emde, & Svejda, 1983) are components, posits that others' emotional communication toward an object informs the observer and guides their perception and behavior (Fischer, 2019; Walle, Reschke, & Knothe, 2017). In such phenomenon, emotion is a key component which helps the learner appraise and reappraise their environment (Fischer, 2019; Walle et al., 2017). However, social appraisal is not merely a case of affective priming (Mumenthaler & Sander, 2012) and with respect to interest for instance, it is likely that the emotional communication of others needs to be referencing the object of interest for social appraisal to occur. Most importantly, affective social learning is about transmission of values and not of knowledge about the world (Fischer, 2019). Moreover, social appraisal learning is an active process in which the learner is actively seeking and processing the affective information from the environment (Walle et al., 2017). Social appraisal can operate automatically: Even if contextual social affective information is sub-optimally perceived, it can still influence emotion recognition of healthy adults (Mumenthaler & Sander, 2015).

In short, the target article insists that we use knowledge representation to learn from others about the external world. We agree and would like to add that we also learn from what others feel. Emotional processes such as social epistemic emotions and affective social learning may play a key role in facilitating the way we learn from the knowledge of others and from the emotions of others. A fascinating research question would be to explore whether processes that rely on affective mechanisms to learn from others are primarily knowledge and/or belief-based.

Financial support. This commentary received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Ainley, M. (2017). Interest: Knowns, unknowns, and basic processes. In P. A. O'Keefe & J. M. Harackiewicz (Eds.), *The science of interest* (pp. 3–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-55509-6_1.

Brun, G., Doguoglu, U, & Kuenzle, D. (2008). In *Epistemology and emotions* (1st ed.). Routledge. https://doi.org/10.4324/9781315580128.

Clément, F., & Dukes, D. (2017). Social appraisal and social referencing: Two components of affective social learning. *Emotion Review*, *9*(3), 253–261. https://doi.org/10.1177/1754073916661634.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Connelly, D. A. (2011). Applying Silvia's model of interest to academic text: Is there a third appraisal? *Learning and Individual Differences*, *21*(5), 624–628. https://doi.org/10.1016/j.lindif.2011.04.007.

Fauth, B., Decristan, J., Decker, A. T., Büttner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and Teacher Education*, *86*, 102882. https://doi.org/10.1016/j.tate.2019.102882.

Fischer, A. (2019). Learning from others' emotions. In D. Dukes & F. Clément (Eds.), *Foundations of affective social learning* (pp. 165–184). Cambridge University Press. https://doi.org/10.1017/9781108661362.008.

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486–496. https://doi.org/10.1016/j.neuron.2014.08.060.

Harris, P. L., Bartz, D. T., & Rowe, M. L. (2017). Young children communicate their ignorance and ask questions. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7884–7891. https://doi.org/10.1073/pnas.1715210114.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x.

Klinnert, M. D., Campos, J. J., Sorce, J. F., Emde, R. N., & Svejda, M. (1983). Emotions as behavior regulators: Social referencing in infancy. In R. Putchik & H. Kellerman (Eds.), *Emotions in early development* (pp. 57–86). Academic Press.

Lau, J. K. L., Ozono, H., Kuratomi, K., Komiya, A., & Murayama, K. (2020). Shared striatal activity in decisions to satisfy curiosity and hunger at the risk of electric shocks. *Nature Human Behaviour*, *4*(5), 531–543. https://doi.org/10.1038/s41562-020-0848-3.

Manstead, A. S. R., & Fischer, A. H. (2001). Social appraisal. In K.R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 221–232). Oxford University Press.

Manstead, A. S. R., Fischer, A. H. (2017). Social referencing and social appraisal: Commentary on the Clément and Dukes (2016) and Walle et al. (2016) articles. *Emotion Review*, *9*(3), 262–263.

Morton, A. (2010). Epistemic emotions. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 385–400). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199235018.003.0018.

Mumenthaler, C., & Sander, D. (2012). Social appraisal influences recognition of emotions. *Journal of Personality and Social Psychology*, *102*(6), 1118.

Mumenthaler, C., & Sander, D. (2015). Automatic integration of social information in emotion recognition. *Journal of Experimental Psychology: General*, *144*(2), 392–399. https://doi.org/10.1037/xge0000059.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In P. Reinhard & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 1–10). Routledge/Taylor & Francis Group.

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (pp. 3–31). American Psychological Association. https://doi.org/10.1037/13274-001.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The epistemically-related emotion scales. *Cognition and Emotion*, *31*(6), 1268–1276. https://doi.org/10.1080/02699931.2016.1204989.

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion (Washington, D.C.)*, *5*(1), 89–102. https://doi.org/10.1037/1528-3542.5.1.89.

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin and Review*, *26*(4), 1377–1387. https://doi.org/10.3758/s13423-019-01598-6.

Walle, E. A., Reschke, P. J., & Knothe, J. M. (2017). Social referencing: Defining and delineating a basic process of emotion. *Emotion Review*, *9*(3), 245–252. https://doi.org/10.1177/1754073916669594.

Knowledge prior to belief: Is extended better than enacted?

Mirko Farina[a] and Andrea Lavazza[b]

[a]Faculty of Humanities and Social Sciences, Universitetskaya St, 1, Innopolis, Republic of Tatarstan 420500, Russian Federation and [b]Centro Universitario Internazionale, Via Antonio Garbasso 42, 52100 Arezzo, AR, Italy.
m.farina@innopolis.ru; http://mirkofarina.weebly.com/
lavazza67@gmail.com; https://www.cui.org/andrea-lavazza/

doi:10.1017/S0140525X2000076X, e152

Abstract

In this commentary, we argue that Phillips et al.'s findings can be used to provide new important insights in the debate between externalists' theories of cognition. In particular, we claim that the results presented in this target article may offer us the conceptual palette needed for a sustained defence of an extended account of cognition over an enactive one.

Phillips et al.'s target article, calls for a shift in focus in theory-of-mind research. More specifically, it proposes a new way to understand theory of mind; one that is – unlike previous versions – deeply grounded on comprehending others' minds in relation to the lived world. This affords the authors to formulate an account of knowledge that is relational and factive in character. In addition, such an account is not reducible to the capacity of attributing true belief and is not modality specific, hence not necessarily innate.

We believe that the empirical findings presented in this target article, pointing out the ontological priority of representations of knowledge over representations of beliefs and the crucial role of the former in facilitating learning from others – can be used to shed light on the debate between externalists theories of cognition in the cognitive sciences. More specifically, we believe it is possible to successfully apply Phillips et al.'s results in the debate between the extended mind thesis (Clark & Chalmers, 1998) and forms of enactivism (such as Noë, 2004; Thompson & Stapleton, 2009).

The extended mind thesis (Clark & Chalmers, 1998; Farina, 2020) is a thesis about human cognition that claims that the cognitive processes that make up our minds can (under specific conditions, the so-called glue and trust conditions) reach beyond the boundaries of individual organisms, so as to include as proper, constitutive aspects of the organism's physical and socio-cultural environment (Kiverstein, Farina, & Clark, 2013). In other words, the extended mind thesis sees the body and the environment (or the technological artefacts located in it with which we reliably interact) as precious – *sometimes* constitutive resources (Farina, 2013; Farina & Levin, In Press) – that we can use in order to enhance our cognitive states.

Research on the extended mind thesis is often said to be arising from functionalist views concerning the "multiple realizability" of cognitive processes and indeed quite a few extended mind theorists (such as Wheeler, 2005) are extremely sympathetic to functionalist and mechanistic accounts of the mind. This means that they believe that mental states are identified by their causal roles and not merely by the medium that realizes them (this understanding is grounded on the so-called parity principle). However, there is also a second strand of research characterizing the extended mind thesis, which is more concerned with the complementarity of inner and outer and so with how internal (neural) and external (extra-neural) resources can work together and eventually become integrated or amalgamated (Rowlands, 2010), so as to form a new, enriched system of cognitive analysis (Menary & Protevi, 2007; Sutton, Harris, Keil, & Barnier, 2010). Crucially, neither of these versions of the extended mind thesis gives up the computational power of our brains nor it repudiates the notion of minimally robust representations (Clark & Toribio, 1994).

On the contrary, enactivism, in all its different strands (see Ward, Silverman, & Villalobos, 2017, for a review) attempts to ground cognition in the biodynamics of living biological systems; hence, it describes cognitive behaviour not only as deeply rooted in our engaged, bodily lives but more profoundly as emerging

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

holistically from the structural coupling and adaptive interplay between the organism and its ecological surroundings (Gallagher & Zahavi, 2020; Noë, 2004; Varela, Thompson, & Rosch, 1991). Consequently, this theory (inspired by phenomenology as well as by Gibson's ecological psychology, 1979) tends to replace the notion of representation with that of sensorimotor contingencies (patterns of contingencies that hold between the movements the perceivers make and what they are able to perceive) and intersubjective affordances (Zahavi, 2002). In addition, in some versions of enactivism (such as Hutto's and Myin's radical enactivism, 2012), this move is accompanied by a contextual progressive repudiation of computationalism. Thus, enactivism (in all its various forms, albeit with different degrees of radicality) asserts that cognitive processes are not content-involving, hence not representation hungry.

We believe that Phillips et al.'s empirical results in showing that representations of knowledge are more basic than representations of beliefs might be used to infer that such a knowledge must be, not only ontologically prior to the other one, but also content-involving and therefore intrinsically representational.

Following the authors' hypothesis that "the primary function of knowledge representation" is "facilitating learning from others" (a claim that might, in principle, be compatible with both extended and enacted accounts), we notice that the former may nevertheless serve this function better than the latter. Contrary to enactivism, which sees cognition as holistically arising from the sensorimotor activity taking place within living biological systems, we claim that the extended mind theory – through is adoption of representations of knowledge and its computationalist/mechanistic roots – provides us with a clearer, more "objective," and perhaps less confused understanding of both the external world in which "others" happen to be, and of the mechanisms that regulate their interactions. This is because in the extended account the elements of a cognitive system that determine the production of knowledge are not meshed indistinguishably together – quite the opposite their respective functions, contributions, and roles can be clearly individuated at any point in time.

Yet we acknowledge that the findings presented in Philips and colleagues' target article cannot be taken to adjudicate the complex debate between these two theories of cognition. This is also because these results are drawn from a very specific and limited domain of inquiry, that of theory of mind. In addition, Phillips et al.'s findings await further replication. Nevertheless, it seems to us that they can be profitably used to suggest an inversion in terms of the debate between extended and enacted.

In summary, we believe that Phillips and colleagues' results may allow to mount as sustained defence of extended over enacted, as the former can describe learning from others about the external world and the related process underlying the production of knowledge within a much clearer theoretical framework, one that is content-involving and does not renounce computation or representations. We believe this is a significant implication of this target article for current debates in empirically informed philosophy of mind.

Conflict of interest

None.

References

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.
Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, *101*(3), 401–431.
Farina, M. (2013). Neither touch nor vision: Sensory substitution as artificial synaesthesia? *Biology & Philosophy*, *28*(4), 639–655.
Farina, M. (2020). Embodiment: dimensions, domains, and applications. *Adaptive Behavior*, *29*(1), 73–99. http://doi.org/10.1177/1059712320912963.
Farina, M., & Levin, S. (In Press). The extended mind thesis: Domains and application. In R. Michael & L. Thomas (Eds.), *Embodied psychology: Thinking, feeling, and acting*. Springer. (accepted)
Gallagher, S., & Zahavi, D. (2020). *The phenomenological mind*. Routledge.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press.
Kiverstein, J., Farina, M., & Clark, A. (2013). *The extended mind thesis*. Oxford University Press. http://www.oxfordbibliographies.com/view/document/obo-9780195396577/obo-97801 95396577-0099.xml
Menary, R., & Protevi, J. (2007). *Cognitive integration: Mind and cognition unbounded*. Palgrave McMillan.
Noë, A. (2004). *Action in perception*. MIT Press.
Rowlands, M. J. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. MIT Press.
Sutton, J., Harris, C. B., Keil, P. G., & Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences*, *9*(4), 521–560.
Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi*, *28*(1), 23–30.
Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi*, *36*(3), 365–375.
Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. A Bradford Book.
Zahavi, D. (2002). First-person thoughts and embodied self-awareness: Some reflections on the relation between recent analytical philosophy and phenomenology. *Phenomenology and the Cognitive Sciences*, *1*, 7–26.

Representation and misrepresentation of knowledge

Mikkel Gerken

University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark.
mikkel@sdu.dk; https://sites.google.com/site/mikkelgerken/

doi:10.1017/S0140525X20001570, e153

Abstract

I argue for three points: First, evidence of the primacy of knowledge representation is not evidence of primacy of knowledge. Second, knowledge-oriented mindreading research should also focus on misrepresentations and biased representations of knowledge. Third, knowledge-oriented mindreading research must confront *the problem of the gold standard* that arises when disagreement about knowledge complicates the interpretation of empirical findings.

The target article by Phillips et al. provides converging evidence for assuming that representations of knowledge are more basic than representations of belief. Although some findings that they take as evidence for representations of knowledge may perhaps be given deflationary interpretations, I am sympathetic to their broad *descriptive* conclusion about the primacy of knowledge representations. Similarly, I agree with their *methodological* conclusion that mindreading research should focus more on knowledge representation. Consequently, I will argue for three further points about knowledge-oriented mindreading research.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

The first point is that it is fallacious to conclude that knowledge itself is primary from evidence that representations of knowledge are primary. Generally, it is fallacious to move from assumptions about the primacy of a *mental representation* to conclusions about the primacy of its *referent*. For example (from Gerken, 2017b), it is a safe bet that representations of water are more basic than representations of hydrogen in terms of ontogenesis, phylogenesis, automatic processing, and so on. But this does not entail that the substance *water* is more basic or primary than *hydrogen*. Such a *representation-representandum* fallacy regarding representations of knowledge may occur in both epistemology and cognitive science. The fallacy is not committed in the target article, although its title – *Knowledge before belief* – might encourage it. Therefore, to ensure that the surveyed evidence is used responsibly, it is important to warn against the *representation-representandum* fallacy. For example, it would be fallacious to take the surveyed evidence to motivate any knowledge-first program concerning knowledge rather than the concept of knowledge, the word "knowledge," and so on (Williamson, 2000).

The second point is that knowledge-oriented mindreading research should study misrepresentations and biased knowledge representations. The primacy of knowledge representations may be partly explained in terms of *bounded rationality*: Representations of knowledge are basic and prominent partly because they are cognitively "cheap" ways of representing complex epistemic matters in a manner that is accurate enough for many purposes. Insofar, as representations of non-factive representations, which require one to keep track of both the mental representation and what it represents, are more cognitively taxing than representations of factive representations, cognitive bounds may partly explain the primacy of knowledge representations. However, bounded cognition involves biases.

The target article provides evidence that knowledge representations are automatically processed (sect. 4.3). But it does not mention that automatic processing of primitive representations often exhibits signature biases (Apperly, 2011; Saxe, 2005). Similarly, misrepresentations of knowledge are not discussed. Given the aim of the target article, its focus on *successful* representations of knowledge is natural. However, this focus suggests a misleading picture of social cognition and overshadows the methodological upshot that knowledge-oriented mindreading research must also focus on misrepresentations and biases. Some research suggests that the patterns of knowledge representations exhibit egocentric bias (Nagel, 2008); focal bias (Gerken, Alexander, Gonnerman, & Waterman, 2020; Gerken & Beebe, 2016); source-content bias (Turri, 2015); subadditivity effects (Dinges, 2018); and so on. Signature biases of knowledge representations are important to study empirically because they are very consequential. For example, they may result in *discriminatory epistemic injustice* which occurs when someone is wronged specifically in her capacity as an epistemic subject (Fricker, 2013, p. 1320; Gerken, 2019). Because representations of knowledge play central roles in navigating social life, cases in which someone is wronged specifically in her capacity as a knower are especially harmful. Generally, given that knowledge representations are central to social cognition, it is important to empirically study their signature biases and the social ramifications thereof (Gerken, 2017a; Spaulding, 2018).

My third point is that knowledge-oriented mindreading research must confront a *problem of the gold standard*: Interpreting empirical data from tasks involving knowledge is often complicated because the gold standard response to the task is disputed. As noted, some researchers argue that particular patterns of knowledge ascriptions reveal a bias (*op. cit.*). But others reject this and argue that these patterns reflect correct responses that illuminate the concept of knowledge, the word "knowledge" or even knowledge itself (e.g., DeRose, 2009; Knobe & Schaffer, 2012; Stanley, 2005). Some even suggest that to explain these patterns of knowledge ascriptions in terms of cognitive bias is to "explain away" the relevant evidence (DeRose, 2009; Stanley, 2005).

Presumably, the false-belief test is a widely employed experimental paradigm partly because of agreement about the gold standard response (e.g., saying that the agent will seek an object where she last saw it rather than at its new location in verbal false-belief tests; Wellman, Cross, & Watson, 2001). In contrast, the gold standard response is *part of what is investigated* in many mindreading tasks involving knowledge. This is simply a methodological challenge for research on knowledge representation and not a reason to stick with established experimental paradigms. Moreover, the problem of the gold standard is far from unique to research on knowledge representation, although it is pertinent to it because many aspects of knowledge are disputed. Minimal properties of knowledge, such as the four considered by Phillips et al. (sect. 2), are good starting points. However, it is disputed how knowledge is related to luck, practical factors, actionability, competence, and so on. Knowledge-oriented mindreading research should study such relationships. But disputes over the gold standard response to tasks involving them constitute a methodological challenge in interpreting findings. Interestingly, the first two points mark specific methodological pitfalls. Given that some patterns of folk knowledge representation are biased (point 2), moving too swiftly from findings about participants' representations of knowledge to conclusions about whether they are correct would exemplify the *representation-representandum* fallacy (point 1).

In sum, here are my three main points:

(1) Evidence of primacy of knowledge representation is not evidence of primacy of knowledge.
(2) Knowledge-oriented mindreading research should also focus on misrepresentations and biased representations of knowledge.
(3) Knowledge-oriented mindreading research must confront *the problem of the gold standard*.

These three points are compatible with the surveyed evidence and the main conclusions that Phillips et al. draw from it. But the points are not included in Phillips et al.'s conclusions. Therefore, I wonder whether they agree with them or not.

Financial support. This study was supported by Danmarks Frie Forskningsfund, Grant Number: 8018-00053B. The author thanks Kenneth Boyd, Uwe Peters, and Shannon Spaulding.

Conflict of interest. None.

References

Apperly, I. A. (2011). *Mindreaders. The cognitive basis of theory of mind*. Psychology Press.
DeRose, K. (2009). *The case for contextualism*. Oxford University Press.
Dinges, A. (2018). Knowledge and availability. *Philosophical Psychology*, *31*(4), 554–573.
Fricker, M. (2013). Epistemic justice as a condition of political freedom? *Synthese*, *190*(7), 1317–1332.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Gerken, M. (2017a). *On folk epistemology. How we think and talk about knowledge*. Oxford University Press.
Gerken, M. (2017b). Against knowledge-first epistemology. In A. Gordon & J. Carter (Eds.), *Knowledge-first approaches in epistemology and mind* (pp. 46–71). Oxford University Press.
Gerken, M. (2019) Pragmatic encroachment and the challenge from epistemic injustice. *Philosophers' Imprint*, *19*(15), 1–19.
Gerken, M., Alexander, J., Gonnerman, C., & Waterman, J. (2020). Salient alternatives in perspective. *Australasian Journal of Philosophy*, *98*(4), 792–810.
Gerken, M., & Beebe, J. R. (2016). Knowledge in and out of contrast. *Noûs*, *50*(1), 133–164.
Nagel, J. (2008). Knowledge ascriptions and the psychological consequences of changing stakes. *Australasian Journal of Philosophy*, *86*, 279–294.
Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences*, *9*(4), 174–179.
Schaffer, J., & Knobe, J. (2012). Contrastive knowledge surveyed. *Noûs*, *46*(4), 675–708.
Spaulding, S. (2018). *How we understand others: Philosophy and social cognition*. Routledge.
Stanley, J. (2005). *Knowledge and practical interests*. Oxford University Press.
Turri, J. (2015). Skeptical appeal: The source-content bias. *Cognitive Science*, *38*(5), 307–324.
Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, *72*(3), 655–684.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Knowledge by default

Robert M. Gordon

Department of Philosophy, University of Missouri, St Louis, MO 63132, USA.
robertmgordon@gmail.com

doi:10.1017/S0140525X20001569, e154

Abstract

The target article presents strong empirical evidence that knowledge is basic. However, it offers an unsatisfactory account of what makes knowledge basic. Some current ideas in cognitive neuroscience – predictive coding and analysis by synthesis – point to a more plausible account that better explains the evidence.

The target article makes a compelling case for an important thesis, the primacy of knowledge attribution. It takes us far in the right direction, veering off only in the final section, where it asks why the capacity for knowledge representation "would have ended up being one that is cognitively basic." The assumption seems to be that either knowledge attribution or belief attribution might have become the basic one, but a certain important function of knowledge attribution (namely, allowing us to learn about the world from others) caused knowledge attribution to "end up" as basic.

I think knowledge attribution is basic in a much more straightforward way. Some current ideas in cognitive neuroscience – predictive coding and analysis by synthesis – show how factual knowledge may be attributed to others simply by default. Withholding or diminishing that attribution, as in attributing a belief that falls short of knowledge, requires additional steps in neural coding and processing. Those extra steps, their added complexity and their drain on resources, suffice to explain the empirical findings: Why some individuals – nonhuman primates, young children, and certain cognitively impaired people – can attribute knowledge but not belief, whereas none attribute belief but not knowledge; and why attributions of knowledge are "more automatic" than those that require additional processing.

1. Predictive coding

As the target article notes, the capacity for knowledge representation is of only limited use in predicting (or in interpreting or explaining) the behavior of others. We can't simply "look at the facts" to predict or explain another's behavior if the other doesn't "share" – that is, doesn't know, isn't aware of – those facts. Nevertheless, it would be folly for the brain to ignore the actual world completely and start over, attempting to build from scratch the set of "facts" that guide the other's behavior. Rather than approach the behavior of others with a blank slate (in Bayesian terms, without priors) – which would be inefficient, if not impossible – the brain very likely implements a predictive strategy. In such a strategy, the actual world – that is, what we ourselves take to be the facts – serves as a starting point, an opening bid or bet, subject to revision ("correction") on the basis of new evidence (Clark, 2013; Koster-Hale & Saxe, 2013).

2. Analysis by synthesis

There is, in fact, a plausible mechanism for implementing such a predictive scheme for anticipating and interpreting others' behavior, as I argue in Gordon (2021). It exploits a strategy that appears to operate in several other areas of cognition, including visual and speech perception: that of analysis by synthesis. Specifically, the brain interprets the behavior of others by testing hypothetical ways of *generating* that behavior. This would involve inverse use of one's own system for planning and generating intentional action: *inverse*, in that what is "given" is the behavior to be generated, and the "result" is whatever best explains this behavior (Baker, Saxe, & Tenenbaum, 2009; Jara-Ettinger, 2019). Such an inversion of the action planning system would run concurrently with its primary "forward" use in generating one's own actions; otherwise, one would have to suspend one's own actions in order to interpret the actions of others. This is consistent with evidence of "motor contagion," or interference effects between observed and executed actions. There appears to be a competition for neural resources, where the same, or strongly overlapping, resources are employed concurrently in goal-directed action planning and in interpreting the goal-directed actions of others (Blakemore & Frith, 2005; Bouquet, Shipley, Capa, & Marshall, 2011).

The inverse use of the planning system for hypothetically generating the actions of others would ordinarily require adjustments of the top-down inputs to the system. These would include adjustments of the factual input, the set of facts that influence planning. In hypothetically generating another's actions, the planning system must be selectively decoupled (disconnected and unplugged) from some of these facts. In the classic "false belief" condition, you see individual A place her treasure at location x. You also see that

(m) the treasure has been moved and is now at a different location y.

If you were planning to steal the treasure, your action planning system would take account of *(m)* and direct you to location y. However, if your system is hypothetically generating A's plan to retrieve A's treasure, the question arises: Does A know about the move? Is A aware that *(m)*? The possibility of attributing ignorance, or not knowing, is simply the possibility of decoupling the action planning system from the fact that *(m)*. (*Egocentric* ignorance acknowledges that there are facts to which our own planning is not yet coupled or connected.) *Knowledge*, on the contrary, is represented simply by nonintervention. That is, one implicitly attributes knowledge that *(m)* simply by *not decoupling*

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

the system from the fact that *(m)*. "Knowledge representations" accordingly consist in nothing more than *access to facts*.

Attributing ignorance consists of decoupling from fact, which is an extra step beyond implicitly attributing knowledge. False belief requires decoupling as well as introducing into the planning process an "as if" fact, such as that the treasure is still at location *x*. True belief for the wrong reason would similarly entail introducing an "as if" fact. (Although it might produce the same actions as the "real" fact, the counterfactual dependencies would differ.) The upshot is that what is really basic is a shared world, where, prior to any corrective processing, everything we ourselves regard as the world, as the facts, is publicly accessible and thus available to others as possible reasons for action.

In sum, the commentary presents strong empirical evidence that knowledge is basic; however, I disagree with the explanation. And I heartily approve the "call to arms" at the end. In my own case it's been revelatory to step outside philosophy and consider possible neural mechanisms that might explain, clarify, and validate the intuitive idea that we understand one another as actors in a shared world.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 329–349.

Blakemore, S.-J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, *43*, 260–267.

Bouquet, C. A., Shipley, T. F., Capa, R. L., & Marshall, P. J. (2011). Motor contagion: Goal-directed actions are more contagious than non-goal-directed actions. *Experimental Psychology*, *58*(1), 71–78.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–253.

Gordon, R. M. (2021). Simulation, predictive coding, and the shared world. In K. Ochsner & M. Gilead (Eds.), *The neural basis of mentalizing* (pp. 36–37). Springer Nature.

Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, *29*, 105–110.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, *79*, 836–848.

Insights into the uniquely human origins of understanding other minds

Tobias Grossmann and Kenn Lacsamana Dela Cruz

Department of Psychology, University of Virginia, Charlottesville, VA 22903, USA.
tg3ny@virginia.edu
kld2db@virginia.edu
https://psychology.as.virginia.edu/people/profile/tg3ny
https://psychology.as.virginia.edu/people/profile/kld2db

doi:10.1017/S0140525X20000710, e155

Abstract

We summarize research and theory to show that, from early in human ontogeny, much information about other minds can be gleaned from reading the eyes. This analysis suggests that eyes serve as uniquely human windows into other minds, which critically extends the target article by drawing attention to what might be considered the neurodevelopmental origins of knowledge attribution in humans.

This commentary complements the target article by drawing attention to a seemingly overlooked line of research and theorizing concerned with what may constitute the origins of knowledge attribution in humans. Specifically, much research has been dedicated to investigate the psychology of perceiving and responding to eye cues as a uniquely human form of social cognition. This research unanimously shows that much information about other minds can be gleaned from eyes and that this information guides social interactions and cooperative decision-making among humans (see Grossmann, 2017, for a review).

First, comparative research suggests that humans, when compared to our closest living primate relatives the chimpanzees, possess a unique sensitivity to eyes and eye cues (Kano & Tomonaga, 2010; Tomasello, Hare, Lehmann, & Call, 2007). This human-specific sensitivity to eyes manifests itself by a more effective and broadened use of eye cues to detect attentional, emotional, and mental-state cues in others. This sensitivity relies upon the unique morphology of the human eyes characterized by their horizontally elongated form and exposed white sclera. Critically, neuroscience research indicates that this sensitivity to eyes is underpinned by a network of brain regions comprising the amygdala at the subcortical level and posterior superior temporal and medial prefrontal cortices at the cortical level, implicated in enhanced attention to eyes and the detection of attentional, emotional, and mental states from eye cues, respectively (Schilbach et al., 2013). Developmental research with infants points to the existence of increased attention to eyes in newborns and indicates that the processes involved in detecting attentional, emotional, and mental states from eyes emerge during the course of the first year of life (Grossmann, 2017). Based on an integration of these findings, the theoretical argument has been advanced that eyes provide a basis for connecting with other minds (Grossmann, 2017). In other words, processes related to privileged orientation and attention to the eyes may help with detecting the presence of other minds, and processes related to decoding information contained in the eye or eye region helps with tracking the contents of other minds, including others' knowledge states.

Second, the existing developmental cognitive neuroscience research with infants, which has not been discussed by Phillips et al. shows that the brain processes underpinning our ability to read other minds develop early in infancy. This is in line with mounting evidence from numerous behavioral studies attesting to infants' mind reading abilities and underlines that behavioral and neuroscience research converges on the notion that access to other minds is an important and early emerging feature of human social cognition. Moreover, the neuroimaging research demonstrates similarity in the brain processes engaged by human infants and adults when processing eye cues (Grossmann, 2017). Based on this similarity, it seems unlikely that different mechanisms (representations) are at play in adults than in infants, further supporting the notion of the early developmental emergence of knowledge attribution. Yet the kind of mental-state understanding attributed to infants based on these neuroscience and behavioral findings does not require the infant to have an explicit (conceptual) grasp of other minds. Indeed, evidence is accumulating that eye-based social cognition might be

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

rooted in automatic processes as demonstrated by neuroscientific research on subliminal face and gaze processing with human infants (Jessen & Grossmann, 2014, 2020).

Third, although the emphasis here was on the ontogenetic and brain origins of human-specific sensitive responding to eyes, eyes are certainly not the only route to understanding other minds. As pointed out by Phillips et al., other sources are also important for informing us about the presence and contents of other minds. For example, there is much research to show that humans use vocal cues in a very similar manner as they use facial and eye cues and sensitivity to voices emerges early in ontogeny (Grossmann, Oberecker, Koch, & Friederici, 2010). Furthermore, in certain contexts, vocal cues have been shown to provide more powerful information regarding another person's mind than facial cues (Schroeder & Epley, 2015). Apart from vocal cues, humans may also rely on information provided through touch (Fairhurst, Löken, & Grossmann, 2014), although when compared to facial and vocal cues, touch as a means for gleaning insights into other minds has been relatively neglected.

From an evolutionary perspective, eye cues are considered to function particularly well during close range interactions without direct physical contact, which are characteristic for many collaborative activities in humans (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). This includes activities such as hunting and gathering (foraging), which is considered the primary and ancestral form of subsistence within the genus *Homo*. Compared to vocal signaling, coordination through eye cues has the advantage that it can occur silently, making it an ideal form of communication between cooperators during group activities such as gathering and hunting, when at the risk of being noticed by predators or prey. Despite the adaptive advantages with respect to human cooperative activities seen in adults, from a developmental perspective, an early emerging sensitivity to eyes might lay the foundation for being able to identify, choose between, and coordinate with cooperative partners.

In summary, we advocate for an approach that systematically and more mechanistically assesses the origins of knowledge about others by taking a developmental cognitive neuroscience perspective in order to advance a more complete understanding of how humans come to understand other minds.

Financial support. T.G. was partly funded by the National Science Foundation #2017229.

Conflict of interest. None.

References

Fairhurst, M. T., Löken, L., & Grossmann, T. (2014). Physiological and behavioral responses reveal 9-month-old infants' sensitivity to pleasant touch. *Psychological Science*, *25*, 1124–1131.

Grossmann, T. (2017). The eyes as windows into other minds: An integrative perspective. *Perspectives on Psychological Science*, *12*, 107–121.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*, 852–858.

Jessen, S., & Grossmann, T. (2014). Unconscious discrimination of social cues from eye whites in infants. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 16208–16213.

Jessen, S., & Grossmann, T. (2020). The developmental origins of subliminal face processing. *Neuroscience & Biobehavioral Reviews*, *116*, 454–460.

Kano, F., & Tomonaga, M. (2010). Face scanning in chimpanzees and humans: Continuity and discontinuity. *Animal Behaviour*, *79*, 227–235.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36* (4), 393–414.

Schroeder, J., & Epley, N. (2015). The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological Science*, *26*, 877–891.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, *52*, 314–320.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*(6), 673–692.

Do knowledge representations facilitate learning under epistemic uncertainty?

Isaac J. Handley-Miner and Liane Young

Department of Psychology, Boston College, Chestnut Hill, MA 02467, USA.
isaac.handley-miner@bc.edu; liane.young@bc.edu; https://moralitylab.bc.edu/

doi:10.1017/S0140525X20001806, e156

Abstract

Phillips and colleagues argue that knowledge representations are more fundamental than belief representations because they better facilitate social learning. We suggest that existing theory of mind paradigms may be ill-equipped to adequately evaluate this claim. Future study should explore learning in situations where there is uncertainty about one's own and others' knowledge, which better mirror real-world social learning contexts.

Phillips and colleagues posit that the adaptive value of "knowledge before belief" is the superiority of knowledge representations for learning in social contexts. Although this hypothesis seems reasonable in the context of paradigms common to theory of mind work, these paradigms eliminate many forms of uncertainty that, in the real world, complicate the process of deciding what to learn, and from whom. In particular, the empirical research featured in the target article leaves little room for (1) uncertainty about the subject's own knowledge and/or (2) uncertainty about other agents' knowledge. Yet, in daily life, people generally experience some degree of uncertainty about these epistemic features. Thus, there is a mismatch between the social learning contexts in these studies and those in the real world.

To illustrate the adaptive learning function of knowledge representations, Phillips and colleagues describe a hypothetical situation in which a ball is placed in one of two boxes in the presence of an agent, and a subject wants to know which box contains the ball. The authors argue that knowledge representations, more so than belief representations, help the subject determine whether they can learn the ball's location from the agent. This situation exemplifies many of the paradigms cited in support of the argument that knowledge representations emerge earlier than belief representations; thus, we will use it to illustrate how these paradigms fail to accommodate various forms of uncertainty that often occur in real-world learning contexts.

The first way in which many theory of mind paradigms eliminate uncertainty is by providing the subject direct observational access to the event of interest (e.g., the ball's location), effectively setting the subject's priors about the event at ceiling (e.g., Bräuer, Call, &

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Tomasello, 2007; Luo & Baillargeon, 2007). In doing so, these paradigms grant the subject the knowledge that the subject represents in the agent; that is, the subject knows where the ball is, and they represent that the agent knows where the ball is. This renders knowledge representations inconsequential for learning; the subject already knows what they would otherwise want to learn.

Second, even in paradigms in which the subject does not have direct observational access to the event of interest, the subject usually has direct observational access to the fact that another agent has direct observational access to the event of interest; that is, the subject sees that the agent sees where the ball is (e.g., Behne, Liszkowski, Carpenter, & Tomasello, 2012; Krachun, Carpenter, Call, & Tomasello, 2009). However, in the real world, the set of situations in which people directly observe another individual acquire complete knowledge without acquiring it themselves is narrow. Often, people are uncertain to some degree about whether someone else has relevant knowledge for their own proximate learning goal. For example, if I want to know whether a fruit is healthy, I may rely on my observation of others eating the fruit. However, I cannot be certain whether the fruit-eaters know that the fruit is healthy, whether they are just very hungry, or whether they simply find the fruit tasty. In other words, I am uncertain about whether they know what I am trying to learn.

Similarly, in addition to eliminating uncertainty around *whether* an agent has knowledge, these paradigms also eliminate uncertainty around *how much* knowledge an agent has. In most paradigms cited in the target article, the agent is either fully ignorant (e.g., has their back turned while the ball is placed in a box) or fully knowledgeable (e.g., can perfectly see which box contains the ball) (e.g., Pratt & Bryant, 1990; Sodian, Thoermer, & Dietrich, 2006). Yet the agents that people seek to learn from are often neither fully knowledgeable nor fully ignorant – they have some amount of relevant knowledge that people must infer from cues such as reputation (e.g., expertise), self-report, or testimony from others.

We believe that a clearer test of Phillips and colleagues' argument about the adaptive value of knowledge representations requires additional empirical investigation into knowledge and belief representations under uncertainty. In situations of uncertainty, do those who can attribute beliefs as well as knowledge still construct knowledge representations with greater automaticity and cognitive ease? Do those who *cannot* attribute beliefs, but can attribute knowledge, still construct knowledge representations? If so, do they act on these knowledge representations as they do in the situations used in existing paradigms? These questions are not an indictment of the knowledge-before-belief claim or the logic of the hypothesis that knowledge representations are more fundamental than belief representations because they better facilitate learning. Rather, we believe that answers to these questions will elucidate how well early knowledge representations actually facilitate social learning, and thus how likely it is that this adaptive argument applies to learning across contexts and across the lifespan.

Although our commentary focuses primarily on two sources of uncertainty underexplored in existing paradigms – uncertainty around one's own and others' knowledge – it is important to note that knowledge representations, on their own, may have limited value for effective social learning absent other mental-state representations. In particular, representing the beliefs, desires, or motivations of others is often critical for helping people to figure out whom to learn from. It is important, for example, to know not only who knows what, but also who can be trusted to share their knowledge, without misleading or obscuring, and without other ulterior motives. Inferring others' beliefs, desires, or motivations could help unlock the adaptive social-learning benefits that the authors argue knowledge representations confer.

In sum, we believe that most existing paradigms examining knowledge attributions in primates and young children do not account for the fact that (1) people are usually learning what they do not already know and (2) people are usually uncertain about what others know and the extent of that knowledge. Under such epistemic uncertainty, do primates and young children still represent knowledge, and, if so, how useful are these representations for learning? Future research that tackles these questions will offer insight into the potential adaptive value of knowledge representations.

Financial support. This study has been supported by the John Templeton Foundation (grant number 61061).

Conflict of interest. None.

References

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, *30*(3), 359–375. https://doi.org/10.1111/j.2044-835X.2011.02043.x.

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*(4), 439–448. https://doi.org/10.1007/s10071-007-0088-1.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x.

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512. https://doi.org/10.1016/j.cognition.2006.10.007.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*(4), 973–982. https://doi.org/10.1111/j.1467-8624.1990.tb02835.x.

Sodian, B., Thoermer, C., & Dietrich, N. (2006). Two- to four-year-old children's differentiation of knowing and guessing in a non-verbal task. *European Journal of Developmental Psychology*, *3*(3), 222–237. https://doi.org/10.1080/17405620500423173.

Three cognitive mechanisms for knowledge tracking

Dora Kampis[a] and Gergely Csibra[b]

[a]Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 Copenhagen, Denmark and [b]Department of Cognitive Science, Central European University, Október 6. u. 7, 1051 Budapest, Hungary.
dk@psy.ku.dk; csibrag@ceu.edu

doi:10.1017/S0140525X20001843, e157

Abstract

We welcome Phillips et al.'s proposal to separate the understanding of "knowledge" from that of "beliefs." We argue that this distinction is best specified at the level of the cognitive mechanisms. Three distinct mechanisms are discussed: tagging one's own representations with those who share the same reality; representing others' representations (metarepresenting knowledge); and attributing dispositions to provide useful information.

In the target article, at least two meanings of "knowledge" are mixed together: "episodic" knowledge (being informed about some state of affairs in the world) and instrumental or semantic competence (being a potential source of information). Although neither of these notions necessarily implies representational mental states (such as beliefs), they are not equal to each other, and should not be expected to be implemented by the same cognitive mechanisms.

With regard to episodic epistemic states, the knowledge–belief distinction may indeed reflect different mechanisms for tracking mental states. Unlike belief attribution, which metarepresents others' mental states, assigning episodic knowledge to others may not require creating a separate representation (Martin & Santos, 2016). An alternative cognitive mechanism is *tagging* one's own representation of reality by symbols of those who have also had access to the state of affairs that gave rise to the representation in question.

Such a tag could be attached to a representation when an episode is co-witnessed with someone (say, agent X), and can be removed when the content of the representation changes in the absence of X. This system can track (1) factive representations (own representations of reality), linking the states of affairs to X only when (2) there is evidence that X has witnessed them (not just happens to believe them) and only until the states remain unchanged, and (3) without preserving the modality that triggered the tagging. Thus, such representations satisfy three of the four criteria for knowledge prescribed by Phillips et al. Furthermore, this tagging mechanism could also explain why "true-belief" attributions are difficult in certain cases: Once X's tag is removed from a representation, it may not be possible to re-attach to it, giving rise to the Gettier-problem (Horschler, Santos, & MacLean, 2019; Kaminski, Call, & Tomasello, 2008). In addition, such a tagging system may also implicitly track altercentric ignorance: A representation that is not tagged by X is knowledge not accessible to X.

However, a cognitive mechanism relying on tagging is unable to generate knowledge representations corresponding to egocentric ignorance (see below for further discussion) and would not be able to account for cases of altercentric interference (produced by a representation attributed to someone else), and aspectuality-based inferences (when the way someone perceives an object leads to representing a different number of entities). We suspect that such phenomena, some of which may be present early in infancy (Kampis & Kovács, 2020; Kovács, Téglás, & Endress, 2010), are to be explained by the same metarepresentational mechanisms that implement belief attribution proper. We agree with Phillips et al. that the representational format underlying many cases of episodic "knowledge" tracking (i.e., tagging) is simpler than what underlies belief attributions. However, the former cannot constitute the basis of the latter because metarepresentations cannot simply emerge from knowledge ascriptions (tagged representations of reality). Rather, tagging may serve as informational input for when attributions of representations become necessary.

Although the tagging system we have outlined above would explain some phenomena of non-belief-based knowledge tracking, it does not support establishing egocentric ignorance, which serves as crucial evidence for Phillips et al.'s conjecture that the main function of knowledge tracking is to promote social learning. However, we believe that the examples of egocentric ignorance listed in the target article either arise from the same attribution system that underlies belief tracking, or represent a completely different notion of knowledge: competence.

This notion comes from the fact that actions of agents carry information about the world: Instrumental actions are adjusted to the environment; communicative actions are designed to produce information. Querying such information sources does not require portraying them as agents possessing episodic knowledge. Instead, an observer may attribute to them a disposition that their instrumental or communicative actions will be informative by reducing the uncertainty of the observer. Such an expectation may be characterized as "egocentric ignorance," yet it entails an entirely different underlying cognitive mechanism from "attributing knowledge" in an episodic sense.

The evidence that Phillips et al. bring forward to support egocentric ignorance in apes is the study by Krachun, Carpenter, Call, and Tomasello (2009), where the subjects inferred the location of the bait from the actions of a competitor. However, the pattern of results suggests that they did so without considering what information was available to the competitor: they simply assumed that he acted competently (cf. Wood et al., 2007). In the same study, human children did indeed attribute episodic knowledge to the competitor when they themselves were ignorant – but they did so also when the competitor had a false belief, suggesting that they relied on a metarepresentational mechanism in tracking the epistemic state of the other. From about 3–4 years of age, children who are ignorant themselves can report on the knowledge of another individual based on their episodic epistemic access (Pillow, 1989; Pratt & Bryant, 1990; Sodian, Thoermer, & Dietrich, 2006; Woolley & Wellman, 1993), most likely reflecting metarepresentational strategies also employed in verbal false-belief tasks.

However, Phillips et al.'s further examples of egocentric ignorance simply require children to portray the putative source of knowledge as being competent to communicate semantic information (e.g., Kovács, Tauzin, Téglás, Gergely, & Csibra, 2014). By default, young children may assume that adults are competent in supplying them with information, but can also fine-tune this assumption on the basis of evidence gathered about potential sources (Begus & Southgate, 2012). When they do so, they do not adjust the amount of "knowledge" attributed to sources, but modulate the sources' expected disposition to produce useful information. This kind of competence attribution indeed promotes learning, but relies on different cognitive mechanisms from those that underlie tracking episodic knowledge. When seeking (or being provided with) information, infants may take the stance that others are competent, but when they provide information to others, they consider the episodic epistemic access of their social partner (Liszkowski, Carpenter, & Tomasello, 2008).

In sum, from the perspective of cognitive mechanisms, knowledge is not "before," but "next to" belief, and it should, in fact, be a plural term.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617.

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*, 72–80.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*, 224–234.
Kampis, D., & Kovács, Á. M. (2020). Seeing the world from others' perspective: 14-month-olds show altercentric modulation effects by others' beliefs. (psyarxiv.com/an7h3).
Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy*, *19*(6), 543–557.
Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibly to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.
Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535.
Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108* (3), 732–739.
Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind?. *Trends in Cognitive Sciences*, *20*(5), 375–382.
Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology*, *47*(1), 116–129.
Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*(4), 973–982.
Sodian, B., Thoermer, C., & Dietrich, N. (2006). Two- to four-year-old children's differentiation of knowing and guessing in a non-verbal task. *European Journal of Developmental Psychology*, *3*(3), 222–237.
Wood, J. N., Glynn, D. D., Phillips, B. C. & Hauser, M. D. (2007). The perception of rational, goal-directed action in nonhuman primates. *Science*, *317*(5843), 1402–1405.
Woolley, J. D., & Wellman, H. M. (1993). Origin and truth: Young children's understanding of imaginary mental representations. *Child Development*, *64*(1), 1–17.

Evolutionary foundations of knowledge and belief attribution in nonhuman primates

Fumihiro Kano[a,b,c] and Josep Call[d]

[a]Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Universitätsstraße 10, 78464, Konstanz, Germany; [b]Max-Planck Institute of Animal Behavior, Am Obstberg 1, 78315, Radolfzell am Bodensee, Germany; [c]Kumamoto Sanctuary, Wildlife Research Center, Kyoto University, Otao 990, Misumi, Uki, Kumamoto, Japan and [d]School of Psychology and Neuroscience, University of St Andrews, KY16 9AJ St Andrews, UK

doi:10.1017/S0140525X20001521, e158

Abstract

Recent findings from anticipatory-looking false-belief tests have shown that nonhuman great apes and macaques anticipate that an agent will go to the location where the agent falsely believed an object to be. Phillips et al.'s claim that nonhuman primates attribute knowledge but not belief should thus be reconsidered. We propose that both knowledge and belief attributions are evolutionary old.

Phillips et al. argued that knowledge attribution is "evolutionary more foundational" than belief attribution, with the former present in monkeys and apes and the latter occurring chiefly in humans. Additionally, Phillips et al. argued that these two theory of mind (ToM) skills are independent from one another. These propositions may seem sensible to comparative psychologists as many (but not all) previous studies in this field have produced evidence for knowledge attribution but not belief attribution in nonhuman animals. However, recent evidence suggests that such a characterization might be too simplistic and overstated. Below, we examine this evidence and its implications by focusing on two meanings of evolutionary foundations of knowledge and belief and conclude that it is conceivable that great apes and macaques have both knowledge and belief representations.

One meaning of "evolutionary more foundational" refers to the temporal emergence of the skills in evolutionary time. Phillips et al. propose that knowledge attribution, which humans share with nonhuman primates, is more ancient than belief attribution, which only humans possess. However, four recent studies with nonhuman primates cast some doubt on this idea (Buttelmann, Buttelmann, Carpenter, Call, & Tomasello, 2017; Hayashi et al., 2020; Kano, Krupenye, Hirata, Tomonaga, & Call, 2019; Krupenye, Kano, Hirata, Call, & Tomasello, 2016). For instance, in the so-called anticipatory-looking false-belief (AL-FB) tests, apes anticipated that an agent will go to the location where the agent falsely believed an object to be. Phillips et al. minimized those findings from AL-FB tests because of the low-level alternatives proposed by some authors, namely that apes "submentalize" (Heyes, 2017) or "see the last location that the agent saw" (Scarf & Ruffman, 2017). However, these alternative explanations were carefully examined and ruled out by subsequent studies with great apes (Kano, Krupenye, Hirata, Call, & Tomasello, 2017; Kano et al., 2019). Notably, Hayashi et al. (2020) recently showed that macaques also pass this AL-FB test, and that inactivation of the macaque medial prefrontal cortex, one of the key regions that support human-adult ToM, disrupted their performance (Hayashi et al., 2020). Thus, Phillips et al.'s claim about belief attribution phylogenetically preceding knowledge attribution based on primate data needs to be reconsidered. Phillips et al.'s misgivings about AL-FB data may be partly motivated by recent replication issues in the AL-FB tests with human infants (Kulke & Rakoczy, 2018) and because the primate studies were inspired by studies with human infants (e.g., Southgate, Senju, & Csibra, 2007), they have also come under scrutiny (Horschler, MacLean, & Santos, 2020). However, the primate work departed from the original versions by introducing several key methodological changes aimed at optimizing the test for nonhuman primates, and the results have been replicated with two different groups of apes and one group of monkeys (Kano, Call, & Krupenye, 2020).

Another meaning of "evolutionary more foundational" refers to one skill being simpler than the other (i.e., less cognitively demanding) and this could explain why in comparative studies is easier to obtain positive results in the knowledge-ignorance than false-belief conditions in traditional nonverbal ToM tests. However, the vast majority of comparative findings come from variations of only two main paradigms: the food-competition in the laboratory (e.g., Kaminski, Call, & Tomasello, 2008) and the violation-of-expectation tests in the wild (e.g., Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011). Although those two paradigms have shown that apes and monkeys attribute knowledge to others, they might not be suitable to detect false-belief attribution because of certain inherent design limitations (as in any single paradigm). How stimuli are presented, how responses are measured as well as the task's motivational substrate can impact performance. For instance, primate violation-of-expectation tests have presented agents performing relatively simple actions, whereas primate AL-FB tests have presented stories in videos illustrating dynamic social interaction between an agent and an antagonist, which may be more intuitively appealing to highly social primates (Kano et al., 2020). Studies that have abandoned the two main paradigms and their inherent limitations have started to produce different results. In fact, in a recent study,

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

apes not only pass false-belief conditions, but also did not find them harder than true-belief or knowledge-ignorance conditions (Buttelmann et al., 2017; also see Buttelmann, Carpenter, & Tomasello, 2009), which casts some doubt on the idea that false-belief conditions are invariably harder than other epistemic conditions. We are not suggesting that AL-FB tests can better capture nonhuman primate ToM in general, but we think it unlikely that a couple of paradigms would be sufficient to capture the full range of socio-cognitive skills that primates deploy in social interaction, particularly knowing that paradigm changes have historically brought substantial empirical and conceptual advances. In keeping with this progress, future studies should investigate whether the notion of aspectuality is part of belief attribution in apes (Low & Watts, 2013).

We liked the authors discussion about the potential functions of knowledge and belief attribution, with the former being particularly useful for learning from others, and the latter for predicting others' behavior, particularly when, unbeknownst to the subjects, the situation has changed. For nonhuman primates, both learning from others and predicting others' behavior should be important, and therefore both knowledge and belief representations can be adaptive. Perhaps for young (human and nonhuman) infants that are dependent on adults, learning from others is more important than predicting others' behavior, but one cannot argue that, for (human and nonhuman) adults, the latter is less important than the former. Imagine, for example, the situation in which an orangutan mother fails to anticipate her child's travel path in a dense forest where visibility is limited; when the child is traveling as usual and did not see (but the mother saw) a branch on which he usually walks was broken. It may be precisely such a situation that critically matters to nonhuman primates – that could happen in their natural lives and affect their fitness. Future studies should endeavor to make the test situations even more ecologically (or ethologically) valid to uncover further elements of belief attribution in primates.

Financial support. FK was supported by Japan Society of Promotion of Science (19H01772 and 20H05000) and JC was supported in part by the European Research Council Synergy Grant 609819 SOMICS.

Conflict of interest. None.

References

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). Great apes distinguish true from false beliefs in an interactive helping task. *PLOS ONE*, *12* (4), e0173793. doi: 10.1371/journal.pone.0173793.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342. doi: 10.1016/j.cognition.2009.05.006.

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., … Hasegawa, I. (2020). Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Reports*, *30*(13), 4433–4444. doi: 10.1016/j.celrep.2020.03.013.

Heyes, C. (2017). Apes submentalise. *Trends in Cognitive Sciences*, *21*(1), 1–2. doi: 10.1016/j.tics.2016.11.006.

Horschler, D. J., MacLean, E. L., & Santos, L. R. (2020). Do non-human primates really represent others' beliefs? *Trends in Cognitive Sciences*, *24*(8), 594–605. doi: 10.1016/j.tics.2020.05.009.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234. doi: 10.1016/j.cognition.2008.08.010.

Kano, F., Call, J., & Krupenye, C. (2020). Primates pass dynamically social anticipatory-looking false-belief tests. *Trends in Cognitive Sciences*, *24*(10), 777–778. doi: 10.1016/j.tics.2020.07.003.

Kano, F., Krupenye, C., Hirata, S., Call, J., & Tomasello, M. (2017). Submentalizing cannot explain belief-based action anticipation in apes. *Trends in Cognitive Sciences*, *21* (9), 633–634. doi: 10.1016/j.tics.2017.06.011.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences*, *116*(42), 20904–20909. doi: 10.1073/pnas.1910095116.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. doi: 10.1126/science.aaf8110.

Kulke, L., & Rakoczy, H. (2018). Implicit theory of mind – An overview of current replications and non-replications. *Data in Brief*, *16*, 101–104. doi: 10.1016/j.dib.2017.11.016.

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*(3), 305–311. doi: 10.1177/0956797612451469.

Marticorena, D. C. W., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science*, *14*(6), 1406–1416. doi: 10.1111/j.1467-7687.2011.01085.x.

Scarf, D., & Ruffman, T. (2017). Great apes' insight into the mind: How great?

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592. doi: 10.1111/j.1467-9280.2007.01944.x.

Belief versus knowledge: An epic battle, but no clear victor

Daniel Heiskell Lassiter

Department of Linguistics, Stanford University, Stanford, CA 94305, USA.
danlassiter@stanford.edu
http://web.stanford.edu/~danlass/

doi:10.1017/S0140525X2000182X, e159

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 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. Although 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 three-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

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

of metarepresentational abilities that would needed to distinguish the child's own model of the world from another's mistaken model, with the result that children under 4 generally mistake false belief for either true belief or no belief. 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, although there is debate as to whether the same theoretical interpretation is appropriate.

The knowledge-first picture is strikingly similar in structure. The parallel three-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 that do not count as knowledge. On this reading, failure to correctly predict actions based on false beliefs is because of 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, although, 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, because 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, and 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 (Oktay-Gür & Rakoczy, 2017, 2020). In a series of experiments, Oktay-Gür and Rakoczy (2017, 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 (although 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.

Conflict of interest

None.

References

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*, 72–80.

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

Oktay-Gür, N., & Rakoczy, H. (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*.

Perner, J. (1991). *Understanding the representational mind*. MIT Press.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

No way around cross-cultural and cross-linguistic epistemology

Edouard Machery[a], H. Clark Barrett[b], and Stephen P. Stich[c]

[a]Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260, USA; [b]Department of Anthropology, University of California, Los Angeles, Los Angeles, CA 90095, USA and [c]Department of Philosophy, Rutgers University, New Brunswick, NJ 08901, USA.
machery@pitt.edu
barrett@anthro.ucla.edu
sstich@ruccs.rutgers.edu
https://www.edouardmachery.com/
http://www.hclarkbarrett.com/

doi:10.1017/S0140525X20001831, e160

Abstract

Phillips and colleagues claim that the capacity to ascribe knowledge is a "basic" capacity, but most studies reporting linguistic

data reviewed by Phillips et al. were conducted in English with American participants – one of more than 6,500 languages currently spoken. We highlight the importance of cross-cultural and cross-linguistic research when one is theorizing about fundamental human representational capacities.

In their fascinating target article, Phillips and colleagues claim that the capacity to ascribe knowledge is a "basic" capacity that does not depend on the capacity to ascribe belief, and they review a large body of evidence in support of this claim: Non-linguistic studies with primates and infants that operationalize the capacity to ascribe knowledge and linguistic studies that examine linguistically how people assign knowledge. On their view, knowledge is conceived by humans, from adults to infants, by apes, and even by monkeys as a factive state, that is not just true belief, that can be obtained on the basis of all sensory modalities and by inference, and that contrasts with ignorance. Although the empirical evidence reviewed by Philips and colleagues is suggestive, it is also flawed, and the goal of this commentary is to highlight its main flaw.

The majority of studies reporting linguistic data reviewed by Phillips et al. were conducted in English with American participants – one of more than 6,500 languages currently spoken. As has been widely discussed in debates about the reliance on WEIRD (western, educated, industrialized, rich, and democratic) participants in psychology (Barrett, 2020; Henrich, Heine, & Norenzayan, 2010; Simons, Shoda, & Lindsay, 2017), there are risks that come from inferring human universality from a small number of possibly unrepresentative cultures and languages. In this case, if there are differences in how knowledge is ascribed across cultures and languages, this poses a challenge for a universalist view of knowledge ascription.

The Geography of Philosophy Project (http://www.geographyofphilosophy.com) is exploring the generalizability of the findings in the linguistic studies discussed by Phillips and colleagues. This project brings together an international team of philosophers, psychologists, linguists, and anthropologists working on five continents to collect data about three important philosophical concepts: the concepts of knowledge, wisdom, and understanding. We study these concepts across a diversity of linguistic, cultural, social, economic, and religious settings.

Some preliminary results confirm the apparent universality of some patterns of knowledge ascription: As mentioned by Phillips and colleagues, across linguistic and cultural settings people tend to deny knowledge in at least some Gettier cases, thus viewing some forms of luck as being incompatible with knowledge (Machery et al., 2017a, 2017b); furthermore, stakes do not matter to the ascription of knowledge (Rose et al., 2019).

But other results, which are directly relevant to the claims made by Phillips and colleagues, are not invariant across cultural and linguistic settings. Phillips and colleagues refer to the finding that English speakers are willing to ascribe knowledge of a given proposition while denying belief in it (Murray, Sytsma, & Livengood, 2013; Myers-Schulz & Schwitzgebel, 2013; replicated with American participants in Kneer, Colaço, Alexander, & Machery, forthcoming), but this dissociation might not be universal: In our preliminary results, we didn't observe the effect in several countries, including Morocco and China. Phillips et al. also take knowledge to be factive: Knowledge that p is only ascribed if the ascriber takes p to be true. They gloss over disagreement among linguists about the factive uses of "know" in English. Although some take factivity to be a semantically required presupposition (e.g., Kiparsky & Kiparsky, 1970), others view it as a pragmatic phenomenon (e.g., Simons, 2007; Vallauri & Masia, 2018). If the latter is true, then it isn't the case that "know" in English expresses the representational capacity Phillips and colleagues have in mind. Be it as it may, we know very little about the factivity of the standard translations of "to know" in the thousands of languages ignored by Phillips and colleagues. Preliminary results suggest much variation in their factive behavior. Although we have observed factive uses in all the languages we have data for, several aspects of factivity vary across languages, including whether factivity is projected through negation, how the factive presupposition is canceled, and whether the terms standardly translated as "know" can be used to express a purely subjective state of confidence.

Non-experimental methods provide further evidence of variation in knowledge ascription. In Chartrand et al. (in prep.), we examine the patterns of colexification of various epistemic lexemes, such as "know" and "understand" in English (see also Georgakopoulos, Grossman, Nikolaev, & Polis, in press). Although "know" and "believe" are often translated by distinct lexemes, in some languages such as Cofán, a single word translates both English expressions. Speakers of these languages might still distinguish the concepts expressed in English by "know" and "believe," but the single lexeme in Cofán that translates both "know" and "believe" may express an altogether different concept.

More generally, the image of knowledge representation and, more generally, of folk epistemology that emerges from our study is at odds with the universalist thrust of Philips and colleagues' article: We observe much variation in the use of "know" and its standard translations and more generally in the use of epistemic vocabulary. For instance, preliminary results suggest variation in whether knowledge is a norm of assertion.

Phillips and colleagues could respond that they are not interested in the meanings of "know" and its translations, but rather in a fundamental representational capacity that may differ from the meanings of these lexical items. However, they "treat knowledge as the ordinary thing meant when people talk about what others do or do not 'know'." Furthermore, if they are not interested in lexical meaning, why do they appeal to the use of "know" in linguistic studies to support their views? Alternatively, they could respond that non-linguistic infants' and primates' studies alleviate the need for cross-linguistic and cross-cultural data: If English speakers and non-linguistic creatures behave similarly, the simplest hypothesis is that all humans and some primates share a common representational capacity. However, simplicity cannot replace direct evidence of universality. Finally, they could respond that concerns about linguistic studies' generalizability leave untouched much of the reviewed evidence, which comes from non-linguistic studies, but without evidence we cannot assume that these results generalize to a diverse sample of human beings.

To conclude, we see no way around painstaking cross-cultural and cross-linguistic research when one is theorizing about fundamental human representational capacities.

Financial support. This project/publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

Conflict of interest. None.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

References

Barrett, H. C. (2020). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences*, *24*, 620–638.

Chartrand, L., Barr, K., Vindrola, F., Allen, C., & Machery, E. (in prep.). Unboxing universality and variation: the distribution of epistemic concepts across culture.

Georgakopoulos, A., Grossman, E., Nikolaev, D., & Polis, S. (in press). Universal and macro-areal patterns in the lexicon. A case-study in the perception-cognition domain. *Linguistic Typology*.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.

Kiparsky, P., & Kiparsky, C. (1970). Fact. In M. Bierwisch & K. E. Heidolph (Eds.), *Progress in linguistics* (pp. 143–173). Mouton.

Kneer, M., Colaço, D., Alexander, J., & Machery, E. (forthcoming). On second thought: Reflection on the reflection defense. In T. Lombrozo, S. Nichols & J. Knobe (Eds.), *Oxford studies in experimental philosophy*. Oxford University Press.

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N. … Hashimoto, T. (2017a). Gettier across cultures. *Nous (Detroit, Mich)*, *51*(3), 645–664.

Machery, E., Stich, S., Rose, D., Alai, M., Angelucci, A., Berniūnas, R. … Zhu, J. (2017b). The Gettier intuition from South America to Asia. *Journal of Indian Council of Philosophical Research*, *34*(3), 517–541.

Murray, D., Sytsma, J., & Livengood, J. (2013). God knows (but does God believe)? *Philosophical Studies*, *166*, 83–107.

Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing that P without believing that P. *Noûs*, *47*(2), 371–384.

Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R. … Zhu, J. (2019). Nothing at stake in knowledge. *Noûs*, *53*(1), 224–247.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua. International Review of General Linguistics. Revue internationale De Linguistique Generale 117*, 1034–1056.

Vallauri, E. L., & Masia, V. (2018). Context and information structure constraints on factivity: The case of know. *Language Sciences*, *66*, 103–115.

Knowledge, belief, and moral psychology

John Mikhail

Georgetown University Law Center, Washington, DC, 20001, USA.
john.mikhail@law.georgetown.edu;
https://www.law.georgetown.edu/faculty/john-mikhail/

doi:10.1017/S0140525X20001788, e161

Abstract

Phillips et al. make a strong case that knowledge representations should play a larger role in cognitive science. Their arguments are reinforced by comparable efforts to place moral knowledge, rather than moral beliefs, at the heart of a naturalistic moral psychology. Conscience, Kant's synthetic a priori, and knowledge attributions in the law all point in a similar direction.

Phillips et al. have produced a fascinating paper, one that makes a strong case that knowledge representations should play a larger role in cognitive science than has occurred until now. As someone whose work in moral psychology has centered on moral knowledge, rather than moral beliefs, for over two decades (e.g., Mikhail, 2000, 2007, 2011, 2014), I am enthusiastic about their project. The questions they are asking, and the interdisciplinary methods they adopt, seem to me exactly the right approach to take to make progress in the theory of mind. I have a few quibbles, but because I am largely persuaded by their main argument, I wish here to accept it substantially as-is and to use this commentary to highlight additional lines of inquiry Phillips et al. might want to consider as they continue to develop this paradigm. All of them reflect the centrality of moral knowledge to moral psychology.

To set the stage, notice first how well moral knowledge fits many of the criteria outlined by the authors for determining whether some representations are more basic than others. For example, consider how natural it is to appeal to moral knowledge in ordinary conversation. We commonly refer to others as *knowing* the difference between right and wrong, rather than believing it. In a similar vein, we refer to others as knowing English, Hindi, Japanese, or other natural languages, rather than believing them. As a complex cognitive capacity, moral knowledge likewise typically emerges early in development (Hamlin, 2013; Kagan & Lamb, 1987), operates largely automatically in adults (Pizzaro & Bloom, 2003), can be preserved in patients who suffer various other cognitive impairments (Nichols, 2004), and is shared to some extent with nonhuman primates (de Waal, 2006; Mikhail, 2014). Unlike many beliefs, we do not forget our moral knowledge (Ryle, 1958), and even a dog knows there is a moral difference between being stumbled over and being kicked (Holmes, 1991/1881).

Three further illustrations of the pivotal role played by moral knowledge in moral psychology seem worth highlighting in this context. Each of them suggests lines of inquiry that Phillips et al. or others may wish to pursue as they seek to deepen this promising research program. First, there is the explicit appeal to "conscience" as a datum of human nature, as manifested in the Universal Declaration of Human Rights and virtually all of the subsequent covenants and treaties that form the modern international law of human rights. What is conscience? The etymology of the original Latin is revealing here: *con-scientia* or "knowledge-with" – characteristically understood as knowing something with another (e.g., God, or one's inner self) (Potts, 1980). Conscience is normally conceived to be a type of knowledge, not belief, and its attributions are historically and culturally ubiquitous. These facts and their behavioral effects may warrant further investigation within the diverse methodological frameworks proposed by Phillips et al.

Second, although they do not discuss this philosophical background, there are strong Kantian overtones to Phillips et al.'s claim that knowledge representations are more basic than belief representations. Indeed, the authors' emphasis on knowledge representations appears at variance with the belief-desire psychology embraced by many contemporary philosophers and psychologists, the main elements of which often derive from a rival philosophical empiricism. For both moral cognition and other forms of cognition, the fundamental epistemological problem for Kant is: How is synthetic a priori knowledge possible? "Synthetic" and its counterpart, "analytic," are adjectives that modify judgments or propositions, whereas "a priori" and its counterpart, "a posteriori," are best understood as adverbs that modify verbs such as "to know" (Wolff, 1973). The key question, for Kant, is thus how synthetic judgments can be *known a priori* (i.e., prior to or independent of experience). There are many difficulties in interpreting Kant and applying his insights to modern cognitive science, of course, but one should not lose sight of the fact that a creative synthesis of Kant and Darwin is possible, in which the former's emphasis on core knowledge representations can be reinterpreted in evolutionary terms (e.g., Lorenz, 1941; Spelke, Lee,

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

& Izard, 2010). The "call to arms" with which Phillips et al. conclude seems to me to lead most naturally in this direction, as does Phillips' other interesting studies on causation, modality, moral judgment, and other topics (e.g., Phillips & Knobe, 2018; Phillips, Morris, & Cushman, 2019).

Finally, many familiar attributions of knowledge and ignorance in legal contexts also lend support to the principal argument advanced by Phillips et al. The clearest example may be the traditional maxim, *ignorantia juris neminem excusat*: "ignorance of the law is no excuse." It reflects what is generally deemed to be an obvious fiction: that everyone is presumed to know the law. In an era in which thousands of obscure statutory or regulatory crimes can serve as the basis of criminal liability, this form of knowledge attribution might seem far-fetched and ridiculous. (For a frequently amusing and sometimes horrifying window into the full catalog of federal crimes, see the @CrimeADay Twitter feed.) A serious and substantive point lies behind the origin of this maxim, however, of which cognitive scientists should take note. Before the advent of modern statutory and regulatory crimes, everyone was presumed to know the law because the law generally reflected customary moral knowledge. Moreover, legally prohibited acts included, or were broadly similar to, those which researchers have recently discovered are condemned by human beings throughout the world, including non-WEIRD (western, educated, industrialized, rich, and democratic) populations in small-scale societies (Barrett et al., 2016; Fessler et al., 2015; Piazza & Sousa, 2016; Saxe, 2016). These prohibitions, in other words, reflect core moral knowledge: a basic grasp of right and wrong, which can be validly denied only by those who fit the legal definition of insanity. Notably, in its most influential formula (the two-prong M'Naughten test), this definition is itself framed in terms of knowledge, rather than belief.

These observations merely scratch the surface of the many interesting possibilities opened up by Phillips et al. I look forward to seeing where their exciting cross-disciplinary research leads.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., … Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *PNAS*, *113*(17), 4688–4693.

de Waal, F. (2006). *Primates and philosophers: How morality evolved*. Princeton University Press.

Fessler, D. M. T., Barrett, H. C., Kanovksy, M., Stich, S., Holbrook, C., Henrich, J., … Laurence, S. (2015). Moral parochialism and contextual contingency across seven societies. *Proceedings of the Royal Society B*, *282*, 1–6.

Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science*, *22* (3), 186–193.

Holmes, Jr., O. W. (1991/1881). *The common law*. New York: Dover.

Kagan, J. and Lamb, S. (Eds.) (1987). *The emergence of morality in young children*. University of Chicago Press.

Lorenz, K. (1941). Kant's doctrine of the a priori in light of contemporary biology. In M. Ruse (Ed.), *Philosophy after Darwin: Classic and contemporary readings* (pp. 231–247). Princeton University Press.

Mikhail, J. (2000). *Rawls's linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in "A theory of justice."* Cornell University PhD.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgement*. Cambridge University Press.

Mikhail, J. (2014). Any animal whatever? Harmful battery and its elements as building blocks of moral cognition. *Ethics*, *124*(4), 750–786.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press.

Phillips, J. & Knobe, J. (2018). The psychological representation of modality. *Mind & Language*, *33*, 65–94.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040.

Piazza, J. & Sousa, P. (2016). When injustice is at stake, moral judgments are not parochial. *Proceedings of the Royal Society B*, *283*, 20152037.

Pizzaro, D. & Bloom, P. (2003). The intelligence of the moral intuitions: Comment on Haidt. *Psychological Review*, *110*, 193–198.

Potts, T. (1980). *Conscience in medieval philosophy*. Cambridge University Press.

Ryle, G. (1958). On forgetting the difference between right and wrong. In A. I. Meldon (Ed.), *Essays in moral philosophy* (pp. 147–159). University of Washington Press.

Saxe, R. (2016). Moral status of accidents. *PNAS*, *113*, 4555–4557.

Spelke, E., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*, *34*(5), 863–884.

Wolff, R. P. (1973). *The autonomy of reason: A commentary on Kant's groundwork of the metaphysics of morals*. New York: Harper.

Knowledge before belief in the history of philosophy

Jessica Moss

Department of Philosophy, New York University, New York, NY 10003, USA.
Jessica.moss@nyu.edu
https://as.nyu.edu/content/nyu-as/as/faculty/jessica-moss.html

doi:10.1017/S0140525X20001612, e162

Abstract

I add support to Phillips et al.'s thesis that representations of knowledge are more basic than representations of belief through a historical account of the development of philosophical theories of knowledge and belief. On the basis of Aristotle's criticisms of his Presocratic predecessors, I argue that Western philosophy developed theories of knowledge long before it developed theories of belief.

To show that representations of knowledge are more basic than representations of belief, Phillips et al. draw on evidence from various branches of psychology, cognitive science, and experimental philosophy. My aim is to add support from a very different source: the history of philosophy. For it turns out that Western philosophy – according, at least, to its first major historian, Aristotle – developed theories of knowledge long before it developed theories of belief.

My evidence is drawn from Aristotle's criticism of his Presocratic predecessors' theories of cognition, in his main psychological treatise. I will show that in this discussion (*De Anima* III.3), Aristotle:

(1) argues that the Presocratics had a theory of knowledge;
(2) argues that they had no theory of false belief; and finally,
(3) introduces as a philosophical innovation a genus of which knowledge and false belief are both species: belief.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

First, Aristotle's Claim 1: the Presocratics had a theory of knowledge.

According to Aristotle, his Presocratic predecessors held that one of the defining features of soul (*psuchê*) is the capacity for what he calls *gnôsis* (*De Anima* 404b9, 404b27-28). The word is sometimes translated as "knowledge," and sometimes as "cognition." Here, the ambiguity is significant. For Aristotle's criticisms amount to the thesis that the Presocratics attempted to give a general account of cognitive activity, but failed precisely because they construed all cognition as knowledge.

On Aristotle's own view there are two broad species of cognition: perception and thought (*to noein*, *to dianoeisthai*). He accuses the Presocratics of conflating the two. Perception is their model for all cognition. Moreover, they construe perception as physical contact between the mind and worldly objects in which the mind comes to resemble the objects. Therefore, on their view, all cognition is true (*DA* 427a21-b3). (For a quick reconstruction of the argument see below; for a detailed reconstruction see Lee, 2005.)

Thus, Aristotle construes the *gnôsis* of the Presocratics as *truth-ensuring contact with reality*.

I submit that this clearly counts as a theory of what we would call knowledge (along the lines of Williamson's "most general factive mental state" (Williamson, 2000)). *Gnôsis* on this account is factive, and, because it involves direct contact and special fit between mind and object, it is more than just true belief. (It also fits Phillips' et al.'s further criteria: it is multi-modal, and allows for representations of egocentric ignorance.) Aristotle is attributing to his predecessors a theory of knowledge.

Second, Aristotle's Claim 2: the Presocratics had no theory of false belief.

Precisely because they construed cognition as they do, Aristotle goes on to argue, his predecessors cannot make sense of cognitive error (427a28-b6). His claim seems to be: If thought is a matter of the mind being made to resemble its object, then if you are thinking at all, you are thinking veridically. Indeed, some of the Presocratics simply deny that false belief exists, arguing that "everything that appears is true" (427b3; cf. *Metaphysics* IV.5, which explicitly equates this slogan with the relativist claim that all opinions are true).

Aristotle is aware that many Presocratics believed in cognitive error. His claim is that they failed to offer a theory of it, or even a theory on which it is possible. In constructing their epistemologies they developed accounts of knowledge, and got stuck there. The clear implication is that it takes a more sophisticated philosopher to develop a theory of false belief. (Compare Plato's criticism of Parmenides in the *Sophist*.)

Finally, Aristotle's Claim 3: the Presocratics had no theory of belief in general.

To account for cognitive error as well as knowledge, Aristotle thinks, we need to recognize a broader category to which both belong. This is precisely what he does in the next part of the discussion, using new technical vocabulary to introduce a new concept.

Thinking, he argues, is composed of two components: *phantasia* (quasi-perceptual appearance), and *hupolêpsis*. *Hupolêpsis* is a genus with several species, some factive and some anti-factive: theoretical knowledge (*epistêmê*), practical knowledge (*phronêsis*), true opinion (true *doxa*), and "the opposites of these" – that is, their false counterparts (427b79-11 and 24–26). Although he does not define *hupolêpsis*, he argues that it presupposes conviction, and suggests that it consists of taking something to be true or false (428a20-428b4). In other words – as many have recognized, and as I argue in detail elsewhere (Moss & Schwab, 2019) – *hupolêpsis* is what modern epistemology calls belief. It is generic taking-to-be-true, which can be true or false, and which when the right conditions are fulfilled constitutes knowledge.

Aristotle does not explicitly accuse the Presocratics of lacking a theory of belief. But he does take their inability to account for cognitive error to show the need for a new theory of thought, one which crucially includes a component so theoretically novel that it requires a neologism ("*hupolêpsis*"). The implication is that his predecessors lacked a theory of belief, and that he is the first to develop one.

Thus, according to Aristotle, in the development of Western philosophy theories of knowledge preceded theories of belief.

I leave to another occasion the question of whether Aristotle is right. A very brief defense: Plato argues that accounting for false belief is a difficult task, and only late and tentatively offers anything like an account of generic belief. (See Moss & Schwab, 2019; for assessment of Aristotle's treatment of Presocratic epistemology, see Lee, 2005.)

At any rate, *if* Aristotle is right, then – granted the plausible assumption that we more easily theorize concepts that are more basic – his account offers further support for Phillips et al.'s contention. For evidently, it comes more easily to humans to construct a philosophical theory of knowledge than one of belief.

Conflict of interest. None.

References

Lee, M. (2005). *Epistemology after Protagoras: Responses to relativism in Plato, Aristotle, and Democritus*. Oxford University Press. https://doi.org/10.1093/0199262225.001.0001.

Moss, J., & Schwab, W. (2019). The birth of belief. *Journal of the History of Philosophy*, *57* (1), 1–32. https://doi.org/10.1353/hph.2019.0000.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

The distinctive character of knowledge

Jennifer Nagel

Department of Philosophy, University of Toronto, Toronto, Canada M5R 2M8. jennifer.nagel@utoronto.ca; http://individual.utoronto.ca/jnagel/

doi:10.1017/S0140525X2000179X, e163

Abstract

Because knowledge entails true belief, it can be hard to explain why a given action is naturally seen as driven by one of these states as opposed to the other. A simpler and more radical characterization of knowledge helps to solve this problem while also shedding some light on what is special about social learning.

Knowing that something is the case is not the same as merely being right about it. The target article offers substantial evidence that knowing is easier to recognize than the state of just having a belief aligned with reality, but we need a sharper picture of knowledge to explain why this is so.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

The authors characterize knowledge through a list of four features, starting with factivity, identified as the condition that "you can only know things that are true." Understood this way, factivity does not distinguish knowledge from true belief, although the next feature stipulates that knowledge is not just true belief. Knowledge has something extra which is lacking in beliefs that are "true by coincidence," but what? The two other listed features of knowledge – that others can know things you don't, and that it is not modality-specific – are unhelpful. Others can have true beliefs you don't, and belief is equally amodal. These four features are supposed to distinguish knowledge from other states, and the article promises to focus on "instances of mental-state representations that have these four signature features." Given that knowledge attribution is supposed to be simpler than belief attribution, mindreaders will presumably not invoke all four features explicitly, conceptualizing the witnessed state as "not just true belief," and so on. But it's unclear what part these features play in any given attribution.

The puzzle deepens when we focus on an ambiguity in mindreading tasks. If some desired object is in a drawer, we expect similar reaching behaviour from the agent who knows that it is there and the agent who just has a true belief. The experimental subject's correct anticipation of that reach does not on its own reveal which state was attributed, if any. Some key studies that the article cites as supportive of easy knowledge attribution actually label that condition as true belief (e.g., Krachun, Carpenter, Call, & Tomasello, 2009). Casually, theorists often gloss the true belief label in terms of knowledge, for example describing the "true belief" condition as consisting of "cases where the [observed agent] knows the peg has moved" (O'Connell & Dunbar, 2003, p. 134). If there is a crucial difference between knowledge and true belief attribution, what explains these patterns of labelling and explaining, and which type of attribution is actually happening?

Theorists are not strictly wrong to label a knowledge condition as true belief, assuming the standard philosophical view that knowledge entails true belief. When someone knows that the peg has moved, it is true both that the peg has moved, and that the agent believes the peg has moved. The target article defends a non-standard view according to which it is possible to know without believing (sect. 5.2), on the basis of intuitive responses to cases in which someone can barely remember a fact. However, the standard view could be retained and the article's overall thesis better supported by a performance error explanation of these borderline cases: They are positively classified by easier processes of knowledge attribution but not by harder processes of belief attribution. Now, even if theorists are right that their knowledge conditions are true belief conditions, this is not to say that experimental subjects are equally indifferent. I agree that it is actually knowledge rather than true belief that is ordinarily attributed in simple control cases, but to defend this position we need to explain the difference in mentalizing.

Here's one proposal: Knowledge is simply a factive mental state, where the factivity condition is read as *necessarily* binding agents only to truths, whereas true belief combines a pair of conditions, one of which (truth) is not mental, and the other of which (belief) is a liberalization of knowledge (Williamson, 2000). We originally track knowledge because the problem of learning what other agents have in mind comes bound with the problem of learning about the larger world; knowledge attribution then works as a special part of the solution to that larger problem. Because other agents have different viewing angles and histories of experience with objects in the shared environment, recognizing signs of their knowledge constitutes a distinctively powerful way of learning about reality. Watching someone who knows where the peg has moved can tell you where the peg is now, assuming we can identify them as knowing, for example through gaze cues. Rather than just being "useful for determining who can accurately inform you about where to look," or as input to calculations about agents' future reliability, factive mentalizing provides a more direct way of learning how things are in the world. Gettiered agents will also inform you accurately, but unlike knowing agents, they are not accurate in virtue of the basic type of mental state they have (belief), so recognizing their mental state does not license a direct updating of one's model of reality. When a cage contains a hidden zebra and a deceptively painted donkey, the agent who sees only the donkey will tell you that there is a zebra in the cage, but that agent's mental relationship to reality is not of a type that necessarily reflects the truth.

Detecting true belief as such, meanwhile, requires separate steps of mentalizing and figuring out what is happening in the world, because belief is defined by relaxing the factivity condition on knowledge. As theoreticians, we are free to execute these steps separately, but we should not assume that experimental participants are doing so. Because belief is a liberalization of knowledge, a wider array of conditions can produce it, so the relevant patterns are harder to learn. Belief-detection patterns are largely derivative of knowledge-detection patterns: for example, in the unwitnessed transfer task, the false belief is that an object is in a location where it was recently known to be (Nagel, 2017). This derivative character of belief attribution explains why some mindreaders attribute knowledge but not belief, whereas none seem capable of attributing belief but not knowledge. The derivative character of belief itself explains why theorists naturally and appropriately explain "true belief" control conditions in terms of knowledge.

Financial support. This study was funded by the Social Sciences and Humanities Research Council of Canada (Insight Grant #502484) and the Schwartz Reisman Institute of the University of Toronto.

Conflict of interest. None.

References

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535.

Nagel, J. (2017). Factive and nonfactive mental state attribution. *Mind & Language*, *32*(5), 525–544.

O'Connell, S., & Dunbar, R. (2003). A test for comprehension of false belief in chimpanzees. *Evolution and Cognition*, *9*(2), 131–140.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Knowing, believing, and acting as if you know

Dilip Ninan

Department of Philosophy, Tufts University, Miner Hall, Medford, MA 02155, USA.
dilip.ninan@tufts.edu; http://www.dilipninan.org

doi:10.1017/S0140525X20001545, e164

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Abstract

Phillips et al. argue that our capacity for representing knowledge is more basic than our capacity for representing belief. But they remain neutral on the further claim that our "belief capacity" depends on our "knowledge capacity." I consider how this further claim might help to explain some of the generalizations the authors catalog, and explore one way of understanding it.

According to Phillips et al., the capacity to represent someone as *knowing* something is more basic than the capacity to represent someone as *believing* something. Their evidence for this claim comes from an impressive variety of sources. They observe, for example, that nonhuman primates attribute knowledge, but not belief, and that the capacity to represent what others know emerges earlier in human development than the capacity to represent what they believe. The claim that knowledge is more basic than belief runs counter to a long-standing tradition in philosophy, cognitive science, and social science that emphasizes *belief* as the primary representational mental state. Think of the false-belief task in psychology, or of the decision theorist's emphasis on beliefs (subjective probabilities) and desires (utilities), or of the traditional epistemologist's attempt to decompose knowledge into belief plus truth plus something else. On the contrary, Phillips et al.'s emphasis on knowledge over belief has an important philosophical counterpart in the recent "knowledge-first" program in epistemology (Williamson, 2000), a connection that will be pursued below.

Let us call the capacity for representing belief the *belief capacity*, and the capacity for representing knowledge the *knowledge capacity*. The view the authors reject has two parts: It says that the belief capacity is more basic than the knowledge capacity, *and* that the knowledge capacity depends on the belief capacity. In arguing that the knowledge capacity is more basic than the belief capacity, the authors reject both parts of this view. But they appear to remain neutral on the further question of whether the converse dependency claim holds (sect. 3.2). Does the belief capacity depend on the knowledge capacity? Or are these two capacities simply independent of each other?

The hypothesis that the belief capacity depends on the knowledge capacity might help to explain some of the generalizations the authors discuss. For example, if you weren't able to represent beliefs without being able to represent knowledge, that would explain why we haven't come across any creatures that can represent belief but not knowledge – we haven't come across any because there couldn't be any. Similarly, the hypothesis that the belief capacity depends on the knowledge capacity would also explain why the belief capacity does not emerge earlier in human development than the knowledge capacity – it doesn't emerge earlier because it couldn't. Of course, other explanations of these facts are possible, but they are likely to be more complex; that provides us with some *prima facie* motivation for exploring the idea that the belief capacity depends on the knowledge capacity.

Let us suppose, for the sake of argument, that this so. How should we understand this dependence? In traditional epistemology, knowledge depends on belief in the sense that knowledge is belief plus truth plus something else. But belief is almost certainly not knowledge plus anything, because one can believe things one does not know (falsehoods being the prime example). Could belief be knowledge *minus* truth? Perhaps, but it is not immediately clear what this would mean (although see Yablo, 2014).

An alternative picture emerges if we focus on certain commonalities between knowledge and belief. As the authors observe, knowledge attributions are sometimes deployed to predict behavior. If you represent Maxi as knowing that the chocolate is in the drawer and represent him as wanting to retrieve the chocolate, you will no doubt expect him to look in the drawer. But, of course, belief attributions can also be deployed in this way: If you represent Maxi as (merely) believing that the chocolate is in the drawer and represent him as wanting to retrieve the chocolate, you will still expect him to look there – and this is so even if you know that the chocolate has actually been moved to the cupboard. One thing this suggests is that when we represent someone as (merely) believing p, we expect them to act as they would if they had known p. When Maxi falsely believes that the chocolate is in the drawer, he acts just as he would if he had known the chocolate was in the drawer. (The "if x had known p" locution here is to be understood in a way that doesn't presuppose that p is in fact true.)

One possibility, then, is that representing someone as believing p involves representing them as acting *as if* they knew p, or perhaps, as being disposed to act as they normally would if they had known p (see also Williamson, 2000, pp. 46–47). If that is (part of) what it is to represent someone as believing something, then it is no wonder that one cannot represent belief if one is unable to represent knowledge. Note also that, according to this proposal, a representation of belief would appear to involve a counterfactual conditional; this might help to account for the apparent link between the ability to engage in (complex) counterfactual reasoning and the ability to pass the false-belief task (e.g., German & Nichols, 2003; Riggs, Peterson, Robinson, & Mitchell, 1998).

Phillips et al. provide some evidence that people are sometimes willing to say that x knows p whereas at the same denying that x believes p (sect. 5.2). This is still possible if belief is understood in the manner suggested above, for one may know p without acting as if one knows p or without being disposed to act as one normally does when one knows p. Indeed, this seems an apt description of the "unconfident examinee" (Myers-Schulz & Schwitzgebel, 2013; Radford, 1966) – she knows the answer, but does not act as if she knows it.

If knowledge is more basic than belief, then it is tempting to think that the belief capacity depends on the knowledge capacity. One form this dependency could take is this: Representing someone as believing involves representing them as acting as if they know. If this proposal is incomplete or entirely wrong-headed, perhaps it will nevertheless serve to provoke others to provide a better account of how belief might depend on knowledge.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science*, *6*(5), 514–523. doi: 10.1111/1467-7687.00309.

Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing that P without believing that P. *Noûs*, *47*(2), 371–384. doi: 10.1111/nous.12022.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Radford, C. (1966). Knowledge: By examples. *Analysis*, *27*(1), 1–11. doi: 10.2307/3326979.
Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, *13*(1), 73–90. doi: 10.1016/s0885-2014(98)90021-1.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press. doi: 10.1093/019925656X.001.0001.
Yablo, S. (2014). *Aboutness*. Princeton University Press. doi: 10.23943/princeton/9780691144955.001.0001.

Beliefs for human-unique social learning

Hilary Richardson

School of Philosophy, Psychology, and Language Sciences, The University of Edinburgh, Edinburgh EH8 9JZ, UK.
hilary.richardson@ed.ac.uk
https://hilaryrichardson.github.io/

doi:10.1017/S0140525X20001600, e165

Abstract

Phillips et al. argue that understanding what others know is central to social cognition across species and that this understanding underlies human-unique accumulation and transmission of cultural knowledge. Knowledge representations can't be both what we have in common with our evolutionary ancestors and what sets us apart from them. Belief representations are necessary for human-unique social learning.

In the course of investigating whether infants and non-human primates represent beliefs, developmental and comparative psychologists generated compelling evidence that these populations represent knowledge. However, evidence that knowledge representations are important for social life across species does not diminish the importance of belief representations for humans. Humans and nonhuman primates – who share the capacity to represent knowledge – nonetheless have dramatically different capacities for accumulating cultural knowledge. Although the capacity to represent knowledge is important for cultural transmission, it is insufficient for human-unique social learning. We must also reason about beliefs.

Phillips et al. argue that knowledge representations – rather than belief representations – support human-unique accumulation and transmission of cultural knowledge because they are clearly in place during infancy and early childhood. In contrast, empirical evidence suggests that social behaviors inherent to human-unique accumulation and transmission of culture develop relatively slowly during early childhood. For example, humans have unique capacities for teaching and learning, aligning perspectives through persuasion, and creating and motivating action toward shared goals. These three social behaviors are honed during early childhood and are supported by mental representations that are tagged with their epistemic history and are not necessarily factive – that is, beliefs.

Reasoning about informants' beliefs enables us to engage in selective social learning and pedagogy: We consider and evaluate the epistemic history of our and others' beliefs to decide who to learn from and what to teach (e.g., Heyes, 2018). As children get older, they increasingly prefer to learn from more accurate informants (Harris & Corriveau, 2011; Harris et al., 2012) and the preference to learn from more accurate – but not physically stronger – informants is predicted by their capacity to reason about diverse beliefs, controlling for age (Brosseau-Liard, Penney, & Poulin-Dubois, 2015). As children reason more flexibly about beliefs, they come to understand that teaching is guided by a teacher's belief about the knowledge gap – rather than the actual knowledge gap – between themselves and their learners (Ziv & Frye, 2004). Theory of mind development in early childhood is linked to children becoming better teachers themselves: Children who pass explicit false-belief tasks selectively present evidence that not only provides knowledge to their learner, but also corrects their learner's particular false belief (controlling for age and numerical conservation reasoning; Bass et al., 2019). Humans do not simply teach and learn to fill gaps in knowledge; we predict, consider, and correct false beliefs.

In addition to correcting others' beliefs, we strategically manipulate them (e.g., Weinstein, 1969). Young children increasingly use the beliefs of the persuadee (e.g., Tricia thinks puppies bite) to tailor their persuasive arguments (e.g., tell Tricia that puppies are gentle rather than quiet; Bartsch, London, & Campbell, 2007). The ability to generate persuasive arguments improves during early childhood and correlates with theory of mind reasoning, controlling for age and language ability (Peterson, Slaughter, & Wellman, 2018; Slaughter, Peterson, & Moore, 2013). Children with disproportionate deficits in theory of mind reasoning show reduced performance on persuasion tasks (Peterson et al., 2018). Skillfully persuading others to adopt our own mental representations requires reasoning about the content and epistemic history of theirs.

Phillips et al. convincingly argue that belief representations are better suited for action prediction than knowledge representations. Accordingly, belief representations also underlie humans' ability to organize and motivate others' actions toward shared goals. False-belief reasoning correlates with production of joint proposals and assignment of roles during pretend play, controlling for age and language abilities (Astington & Jenkins, 1995) and 6-year-old children use first- and second-order belief representations to coordinate on tasks with their peers (Grueneisen, Wyman, & Tomasello, 2015).

Regardless of exactly when children or infants begin to represent others' beliefs (e.g., Poulin-Dubois et al., 2018), there is ample evidence that belief representations are used more flexibly and in service of increasingly sophisticated social behaviors – including social behaviors inherent to human-unique accumulation and transmission of knowledge – during early childhood. This continued development reflects genuine conceptual change in theory of mind representations, rather than a gradual unmasking of competence as language and executive functions improve. Childhood theory of mind reasoning is predicted by earlier theory of mind capacities over and above these other skills (Peterson & Wellman, 2019; Richardson et al., unpublished data; Wellman, Fang, & Peterson, 2011) and is mirrored by continued development in brain regions that support social cognition (Richardson, Lisandrelli, Riobueno-Naylor, & Saxe, 2018), including specialization of the right temporoparietal junction for reasoning about mental states (beliefs, desires, and emotions; Richardson et al., 2020). The capacity to reason about beliefs – deliberately, with slow and gradual improvement during childhood, and with consequences for populations for whom this is challenging – is intrinsic to human-unique accumulation and transmission of cultural knowledge.

Acknowledging a relatively slower developmental trajectory for human-unique accumulation and transmission of cultural

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

knowledge additionally allows formal education to play a role. In many societies, school provides a venue for children to learn not only how to read and write, but also how to become a citizen of their community and culture (Zigler & Trickett, 1978). Explicit false-belief reasoning in early childhood predicts school readiness (controlling for age, language, IQ, attention shifting, and executive functions, Blair & Razza, 2007; for a review, see Astington & Pelletier, 2005), suggesting that, in addition to enabling increasingly sophisticated social behaviors, ongoing theory of mind development enables children to capitalize on institutions specifically in place for the accumulation and transmission of cultural knowledge.

As Phillips et al. propose, one important goal for future theory of mind research is to offer a description of early developing, evolutionarily shared, automatic, and preserved capacities – like knowledge representations – and the social behaviors that they can and cannot support. A second and equally important goal is to offer a description of ongoing conceptual change in childhood – which includes the development of theory of mind capacities that are core to human-unique intelligence and culture.

Acknowledgments. I am grateful to Ashley Thomas, Shari Liu, and Lindsey Powell for feedback, and to Rebecca Saxe for encouragement.

Conflict of interest. None.

References

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition & Emotion*, *9*(2–3), 151–165.

Astington, J. W., & Pelletier, J. (2005). Theory of mind, language, and learning in the early years: Developmental origins of school readiness. In B. D. Homer & C. Tamis-Lemonda (Eds.), *The development of social cognition and communication* (pp. 205–230). Lawrence Erlbaum.

Bartsch, K., London, K., & Campbell, M. D. (2007). Children's attention to beliefs in interactive persuasion tasks. *Developmental Psychology*, *43*(1), 111.

Bass, I., Gopnik, A., Hanson, M., Ramarajan, D., Shafto, P., Wellman, H., & Bonawitz, E. (2019). Children's developing theory of mind and pedagogical evidence selection. *Developmental Psychology*, *55*(2), 286.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663.

Brosseau-Liard, P., Penney, D., & Poulin-Dubois, D. (2015). Theory of mind selectively predicts preschoolers' knowledge-based selective word learning. *British Journal of Developmental Psychology*, *33*(4), 464–475.

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). "I know you don't know I know…" Children use second-order false-belief reasoning for peer coordination. *Child Development*, *86*(1), 287–293.

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1179–1187.

Harris, P. L., Corriveau, K. H., Pasquini, E. S., Koenig, M., Fusaro, M., & Clément, F. (2012). Credulity and the development of selective trust in early childhood. In M. Beran, J. L. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 193–210). Oxford University Press.

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Peterson, C. C., Slaughter, V., & Wellman, H. M. (2018). Nimble negotiators: How theory of mind (ToM) interconnects with persuasion skills in children with and without ToM delay. *Developmental Psychology*, *54*(3), 494.

Peterson, C. C., & Wellman, H. M. (2019). Longitudinal theory of mind (ToM) development from preschool to adolescence with and without ToM delay. *Child Development*, *90*(6), 1917–1934.

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., … Low, J. (2018). Do infants understand false beliefs? We don't know yet – A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, *48*, 302–315.

Richardson, H., Koster-Hale, J., Caselli, N., Magid, R., Benedict, R., Olson, H., … Saxe, R. (2020). Reduced neural selectivity for mental states in deaf children with delayed exposure to sign language. *Nature Communications*, *11*(1), 1–13.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*(1), 1027.

Slaughter, V., Peterson, C. C., & Moore, C. (2013). I can talk you into it: Theory of mind and persuasion behavior in young children. *Developmental Psychology*, *49*(2), 227.

Weinstein, E. A. (1969). The development of interpersonal competence. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 753–775). Rand McNally.

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, *82*(3), 780–792.

Zigler, E., & Trickett, P. K. (1978). IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologist*, *33*(9), 789.

Ziv, M., & Frye, D. (2004). Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development*, *19*(4), 457–477.

Semantic memory before episodic memory: How memory research can inform knowledge and belief representations

R. Shayna Rosenbaum, Julia G. Halilova, and Thanujeni Pathman

Department of Psychology and Centre for Vision Research, York University, Toronto, ON M3J 1P3, Canada.
shaynar@yorku.ca, gkhalilova@gmail.com, tpathman@yorku.ca

doi:10.1017/S0140525X20001867, e166

Abstract

Knowledge and belief attribution are discussed in the context of episodic and semantic memory theory and research, with reference to patient-lesion and developmental studies under naturalistic conditions. Consideration of how episodic and semantic memory relate to each other and intersect in the real world, including how they fail, can illuminate the approach to studying how people represent others' minds.

Phillips and colleagues propose that the ability to represent knowledge is separate from, and fundamental to, representing beliefs. We believe that this account lays the groundwork for future studies on theory of mind (ToM) but would benefit from consideration of memory theory and research. We suggest turning to semantic and episodic memory, not only because these forms of memory resemble, and potentially contribute to, knowledge and belief representations, but also because conceptual advances in understanding semantic versus episodic memory involved assessing the directionality of their relationship, areas of overlap, and how they are expressed in the real world. Addressing similar issues in relation to knowledge and belief representations could lead to further progress in understanding our capacity to represent others' minds.

Episodic memory is memory for detailed personal experiences that occurred at a specific time and place (Tulving, 1972, 1983). By contrast, *semantic memory* contains knowledge about the world and oneself in a context-free form that can be represented separately from the specific experiences in which the knowledge was acquired. In proposing his theory on episodic memory, Endel Tulving was cognizant of similarities with semantic memory but needed to demonstrate how episodic memory is distinct (Renoult & Rugg, 2020). Similar to belief representations, episodic memory is experience-dependent and late-developing, both

phylogenetically and ontogenetically. It is flexible and reconstructive, so that retrieval from episodic memory lays down new, unique episodic memories. Retrieval of knowledge from semantic memory should leave its content unaltered, although it provides content for new episodic memories. In this way, semantic memory represents an essential foundation for episodic memory: without semantic memory, the emergence of episodic memory should not be possible (Tulving, 1983).

Findings in child development are consistent with the proposed relationship between semantic and episodic memory. Semantic memory typically emerges before episodic memory (Nelson & Fivush, 2004; Robertson & Köhler, 2007). Individuals with developmental amnesia because of early compromise of the hippocampal memory system experience selective deficits in acquiring episodic memories but can acquire personal and general knowledge (Rosenbaum et al., 2011; Vargha-Khadem et al., 1997). The idea that semantic memory is a prerequisite for episodic memory nevertheless has been challenged on the grounds that older adults with semantic dementia (temporal-variant frontotemporal dementia) show an opposite dissociation to that seen in developmental amnesia: impaired semantic memory with relatively intact episodic memory (Viard et al., 2013; Westmacott, Leach, Freedman, & Moscovitch, 2001). Moreover, in neurotypical populations, memory for specific experiences is the route by which at least some knowledge is acquired or extracted (Sommer, 2017; Yee, Chrysikou, & Thompson-Schill, 2013).

Even if semantic memory is fundamental to episodic memory, there is widespread agreement that they are intertwined, especially in the real world (Prebble, Addis, & Tippett, 2013; Renoult, Irish, Moscovitch, & Rugg, 2019). For instance, the self, once viewed as a defining feature of episodic memory (Tulving, 1985), is recognized as an essential part of personal semantic memory (i.e., memory for self-related facts; Grilli & Verfaellie, 2014). In semantic dementia, concepts that are personally significant (e.g., patient's own coffee mug) are less vulnerable to disruption than concepts that do not hold such meaning (e.g., doctor's coffee mug; Snowden, Griffiths, & Neary, 1996). More ecologically valid measures of semantic memory have revealed further areas of overlap, with deficits observed in patients with hippocampal damage on tasks involving construction of narratives (Rosenbaum, Gilboa, Levine, Winocur, & Moscovitch, 2009; Verfaellie, Bousquet, & Keane, 2014) and scenes (Lynch, Keane, & Verfaellie, 2020). This is not surprising given marked differences between lab-based, list-learning tests and real-world autobiographical tests of episodic memory in terms of their neural basis (Gilboa, 2004; McDermott, Szpunar, & Christ, 2009) and developmental trajectory (Pathman, Samson, Dugas, Cabeza, & Bauer, 2011).

The methods used to assess knowledge versus belief attribution may similarly fall short of simulating reality, where the divide between knowledge and belief is not always so clear. It may be difficult to verify that "Suzy knows where to buy an Italian newspaper" if this scenario from the target article was set in the real world. Correspondence of lab-based tests to real-world situations has also been questioned in studies examining the relationship between episodic memory and ToM. These investigations were prompted by simulation theories that view the ability to recollect one's own past mental states as central to inferring the current mental states of others (Buckner & Carroll, 2007; Shanton & Goldman, 2010). Evidence that episodic memory and ToM emerge close in time in early development (Perner & Ruffman, 1995), are impaired together in autism spectrum disorder (Ciaramelli et al., 2018), and rely on an overlapping neural substrate (Spreng, Mar, & Kim, 2009) supports this idea. Contrary to this view, however, patients with severely impaired episodic memory because of hippocampal dysfunction perform at the same level as controls on a large number of tests known to be sensitive to ToM (Rabin, Braverman, Gilboa, Stuss, & Rosenbaum, 2012; Rosenbaum, Stuss, Levine, & Tulving, 2007), including the kinds of measures described in the target article that assess the capacity to represent others' knowledge versus beliefs (cf. Krachun, Carpenter, Call, & Tomasello, 2009; Stuss, Gallup, & Alexander, 2001). The findings seem at variance with the idea that ToM depends on episodic memory, but other studies in amnesia suggest that when familiar others are concerned, episodic memory is required for ToM (Rabin, Carson, Gilboa, Stuss, & Rosenbaum, 2013; Rabin, Carson, Gilboa, Stuss, & Rosenbaum, 2016). This research highlights another lesson that may be gleaned from research on semantic versus episodic memory: Extending the special population method to include patients with focal brain lesions to infer the causal structure of knowledge and belief attributions (Rosenbaum, Gilboa, & Moscovitch, 2014).

In sum, we describe several issues that have been raised in the study of semantic versus episodic memory that are relevant to knowledge versus belief attribution and how they have been addressed. A growing number of memory researchers have adopted a more fluid, process-oriented view of semantic and episodic memory in place of the more traditional systems approach. Although a consensus has not yet been reached, the process of re-evaluating these forms of memory has encouraged more rigorous testing of causal claims under naturalistic conditions. Future research on the relationship between knowledge and belief attribution might benefit from a similar approach.

Financial support. This commentary was supported by the Canada First Research Excellence Fund (CFREF) Vision: Science to Applications (VISTA) Program and a York Research Chair in Cognitive Neuroscience of Memory (RSR).

Conflict of interest. None.

References

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*, 49–57. https://doi.org/10.1016/j.tics.2006.11.004.

Ciaramelli, E., Spoglianti, S., Bertossi, E., Generali, N., Telarucci, F., Tancredi, R., … Igliozzi, R. (2018). Construction of past and future events in children and adolescents with ASD: Role of self-relatedness and relevance to decision-making. *Journal of Autism and Developmental Disorders*, *48*, 2995–3009. https://doi.org/10.1007/s10803-018-3577-y.

Gilboa, A. (2004). Autobiographical and episodic memory – one and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia*, *42*, 1336–1349. https://doi.org/10.1016/j.neuropsychologia.2004.02.014.

Grilli, M. D., & Verfaellie, M. (2014). Personal semantic memory: Insights from neuropsychological research on amnesia. *Neuropsychologia*, *61*, 56–64. https://doi.org/10.1016/j.neuropsychologia.2014.06.012.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*, 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x.

Lynch, K., Keane, M. M., & Verfaellie, M. (2020). The status of semantic memory in medial temporal lobe amnesia varies with demands on scene construction. *Cortex*, *131*, 114–122. https://doi.org/10.1016/j.cortex.2020.07.005.

McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, *47*, 2290–2298. https://doi.org/10.1016/j.neuropsychologia.2008.12.025.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review*, *111*, 486–511.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Pathman, T., Samson, Z., Dugas, K., Cabeza, R., & Bauer, P. J. (2011). A "snapshot" of declarative memory: Differing developmental trajectories in episodic and autobiographical memory. *Memory (Hove, England)*, *19*, 825–835. https://doi.org/10.1080/09658211.2011.613839.

Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, *59*, 516–548. https://doi.org/10.1006/jecp.1995.1024.

Prebble, S. C., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin*, *139*, 815–840. https://doi.org/10.1037/a0030146.

Rabin, J. S., Braverman, A., Gilboa, A., Stuss, D. T., & Rosenbaum, R. S. (2012). Theory of mind development can withstand compromised episodic memory development. *Neuropsychologia*, *50*, 3781–3785. https://doi.org/10.1016/j.neuropsychologia.2012.10.016.

Rabin, J. S., Carson, N., Gilboa, A., Stuss, D. T., & Rosenbaum, R. S. (2013). Imagining other people's experiences in a person with impaired episodic memory: The role of personal familiarity. *Frontiers in Psychology*, *3*, 588. https://doi.org/10.3389/fpsyg.2012.00588.

Rabin, J. S., Olsen, R. K., Gilboa, A., Buchsbaum, B. R., & Rosenbaum, R. S. (2016). Using fMRI to understand event construction in developmental amnesia. *Neuropsychologia*, *90*, 261–273. https://doi.org/10.1016/j.neuropsychologia.2016.07.036.

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: The semantic-episodic distinction. *Trends in Cognitive Sciences*, *23*, 1041–1057. https://doi.org/10.1016/j.tics.2019.09.008.

Renoult, L., & Rugg, M. D. (2020). An historical perspective on Endel Tulving's episodic-semantic distinction. *Neuropsychologia*, *139*, 107366. https://doi.org/10.1016/j.neuropsychologia.2020.107366.

Robertson, E. K., & Köhler, S. (2007). Insights from child development on the relationship between episodic and semantic memory. *Neuropsychologia*, *45*, 3178–3189. https://doi.org/10.1016/j.neuropsychologia.2007.06.021.

Rosenbaum, R. S., Carson, N., Abraham, N., Bowles, B., Kwan, D., Köhler, S., … Richards, B. (2011). Impaired event memory and recollection in a case of developmental amnesia. *Neurocase*, *17*, 394–409. https://doi.org/10.1080/13554794.2010.532138.

Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia*, *47*, 2181–2187. https://doi.org/10.1016/j.neuropsychologia.2008.11.028.

Rosenbaum, R. S., Gilboa, A., & Moscovitch, M. (2014). Case studies continue to illuminate the cognitive neuroscience of memory. *Annals of the New York Academy of Sciences*, *1316*, 105–133. https://doi.org/10.1111/nyas.12467.

Rosenbaum, R. S., Stuss, D. T., Levine, B., & Tulving, E. (2007). Theory of mind is independent of episodic memory. *Science*, *318*, 1257. https://doi.org/10.1126/science.1148763.

Shanton, K., & Goldman, A. (2010). Simulation theory. Wiley interdisciplinary reviews. *Cognitive Science*, *1*, 527–538. https://doi.org/10.1002/wcs.33.

Snowden, J. S., Griffiths, H. L., & Neary, D. (1996). Semantic-episodic memory – interactions in semantic dementia: Implications for retrograde memory function. *Cognitive Neuropsychology*, *13*, 1101–1137.

Sommer, T. (2017). The emergence of knowledge and how it supports the memory for novel related information. *Cerebral Cortex*, *27*, 1906–1921. https://doi.org/10.1093/cercor/bhw031.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*, 489–510. https://doi.org/10.1162/jocn.2008.21029.

Stuss, D. T., Gallup, G. G. Jr, & Alexander, M. P. (2001). The frontal lobes are necessary for "theory of mind." *Brain*, *124*, 279–286. https://doi.org/10.1093/brain/124.2.279.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donalson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Clarendon Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology 26*:1–12.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*, 376–380. https://doi.org/10.1126/science.277.5324.376.

Verfaellie, M., Bousquet, K., & Keane, M. M. (2014). Medial temporal and neocortical contributions to remote memory for semantic narratives: Evidence from amnesia. *Neuropsychologia*, *61*, 105–112. https://doi.org/10.1016/j.neuropsychologia.2014.06.018.

Viard, A., Desgranges, B., Matuszewski, V., Lebreton, K., Belliard, S., de La Sayette, V., … Piolino, P. (2013). Autobiographical memory in semantic dementia: New insights from two patients using fMRI. *Neuropsychologia*, *51*, 2620–2632. https://doi.org/10.1016/j.neuropsychologia.2013.08.007.

Westmacott, R., Leach, L., Freedman, M., & Moscovitch, M. (2001). Different patterns of autobiographical memory loss in semantic dementia and medial temporal lobe amnesia: A challenge to consolidation theory. *Neurocase*, *7*, 37–55. https://doi.org/10.1093/neucas/7.1.37.

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2013). Semantic memory. In K. Ochsner & S. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience, volume 1: Core topics* (pp. 353–374). Oxford University Press.

Ignorance matters

Amanda Royka and Julian Jara-Ettinger

Department of Psychology, Yale University, New Haven, CT 06520-8205, USA.
amanda.royka@yale.edu; julian.jara-ettinger@yale.edu;
https://compdevlab.yale.edu/

doi:10.1017/S0140525X20001636, e167

Abstract

The ability to reason about ignorance is an important and often overlooked representational capacity. Phillips and colleagues assume that knowledge representations are inevitably accompanied by ignorance representations. We argue that this is not necessarily the case, as agents who can reason about knowledge often fail on ignorance tasks, suggesting that ignorance should be studied as a separate representational capacity.

How do we reason about agents who are ignorant? When we interact with someone that has partial or incomplete knowledge, we can flexibly understand and predict their behavior depending on whether their ignorance is easy to remedy (what's inside this box?), out of their control (will it rain today?), or irrelevant to their goals (do we have free will?). Similarly, when we recognize that we don't know something, we can rectify our ignorance through exploration or by watching more knowledgeable agents act.

Phillips et al. make a compelling case that, within theory of mind, knowledge is a more basic representation than belief. But, in doing so, Phillips et al. also treat knowledge and ignorance as two sides of the same representational coin. However, the representational demands of knowledge and ignorance are not necessarily equivalent. For instance, one of the simplest ways to represent ignorance would be as the absence of knowledge. This, however, would require a negation-like representation of a knowledge state. Because the ability to apply negation over mental representations appears to be absent in younger children (Feiman, Mody, Sanborn, & Carey, 2017; Mody & Carey, 2016; Nordmeyer & Frank, 2014; Reuter, Feiman, & Snedeker, 2018) and is weak in nonhuman primates (Call & Carpenter, 2001), even this simple relationship would already predict that representations of ignorance are not an inevitable consequence of representations of knowledge.

Even if children and nonhuman primates could represent ignorance as a consequence of their ability to represent knowledge, this alone would not provide the computations needed to predict and understand the behavior of ignorant agents, making these representations of limited use. Indeed, predicting the behavior of an ignorant agent goes far beyond merely expecting that they will not act in a knowledgeable way: Accurate predictions about ignorant agents involve determining whether they will

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

choose to gather information, and how they will act to maximize their chance of success under uncertainty.

Importantly, these concerns do not simply reflect theoretical questions about the nature of ignorance representations. The few empirical studies that test for an intuitive theory of ignorance suggest that these representations have a tenuous correlation with knowledge representations. Although some sensitivity to ignorance appears early in development (Koenig & Echols, 2003; O'Neill, 1996), children's understanding of ignorance continues to develop after children have a mature understanding of knowledge. Young children exhibit egocentric errors, attributing their own knowledge to ignorant agents (Birch & Bloom, 2003; Hogrefe, Wimmer, & Perner, 1986; Mossler, Marvin, & Greenberg, 1976; Sullivan & Winner, 1991; Wellman & Liu, 2004); they fail to predict that agents searching for a hidden object will choose randomly (Friedman & Petrashek, 2009; Ruffman, 1996); and they do not expect ignorant agents to seek additional information when necessary (Huang, Hu, & Shao, 2019).

Similarly, there is little evidence that nonhuman primates can predict the actions of ignorant agents (Drayton & Santos, 2018; Horschler, Santos, & MacLean, 2019; Karg, Schmelz, Call, & Tomasello, 2015b; Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011; Martin & Santos, 2016). Many experiments examining nonhuman primate theory of mind directly contrast knowledge and ignorance in a single task, which means that subjects can succeed by (1) only representing knowledge, (2) only representing ignorance, or (3) representing both (e.g., Flombaum & Santos, 2005; Hare, Call, Agnetta, & Tomasello, 2000; Karg, Schmelz, Call, & Tomasello, 2015a), making it impossible to discern which representations are guiding subjects' behavior. Even looking-time tasks that probe knowledge and ignorance under different conditions do not provide clear evidence of ignorance representations. For example, after seeing an object hidden in one of two boxes, rhesus macaque monkeys look equally long at the display when an ignorant demonstrator reaches for the correct or incorrect box (Drayton & Santos, 2018; Marticorena et al., 2011). Crucially, these results are consistent with two competing explanations: Subjects may be unsurprised because both actions are consistent with their prediction that the ignorant agent will search randomly or they may be unsurprised because they made no prediction at all. The former is consistent with Phillips et al.'s proposal that nonhuman primates are able to make predictions about both knowledgeable and ignorant agents. However, the latter would suggest that rhesus macaques either cannot represent ignorance or cannot form predictions about ignorant agents, despite having expectations about the behavior of knowledgeable agents (Drayton & Santos, 2018; Marticorena et al., 2011). Similar concerns also apply to "ignorance" conditions in looking-time studies with infants (Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; Luo & Johnson, 2009).

Taken together, these empirical findings suggest that children and nonhuman primates may not have a rich understanding of ignorance despite being able to successfully reason about knowledgeable agents. This presents an exciting opportunity to reevaluate the common assumption that ignorance representation inevitably accompanies knowledge representation. One possibility is that knowledge is a primary representation out of which ignorance representations are later derived – through negation or otherwise. Such a relationship would explain the developmental lag in ignorance understanding in children and make testable predictions about the status of ignorance representations in nonhuman primates depending on the hypothesized requirements to build this secondary representation. Alternatively, knowledge and ignorance representations may be independent from one another, combining later in life to support reasoning about agents with partial or incomplete knowledge. Critically, in either case, these proposals are consistent with Phillips et al.'s view of the primacy of knowledge representations.

Or perhaps, Phillips et al. are right: Knowledge and ignorance representations may be impossible to disentangle, developmentally indistinguishable (with previous ignorance failures representing only task demands), and best understood in tandem. The task is now to clearly articulate this relationship and design empirical investigations of ignorance representations in their own right, rather than as a control condition for studies of knowledge. A complete account of mental-state representations must explain how ignorance is derived, what (if any) additional representational machinery is necessary, and whether the hypothesized relationship predicts any critical gaps in development of representations of knowledge and ignorance. The answers to these questions are essential not only for understanding this representational capacity, but also for understanding our knowledge representation system and our ability to interpret and predict epistemic actions.

Conflict of interest. None.

References

Birch, S. A., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, *14*(3), 283–286.

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, *3*(4), 207–220.

Drayton, L. A., & Santos, L. R. (2018). What do monkeys know about others' knowledge? *Cognition*, *170*, 201–208.

Feiman, R., Mody, S., Sanborn, S., & Carey, S. (2017). What do you mean, no? Toddlers' comprehension of logical "no" and "not." *Language Learning and Development*, *13*(4), 430–450.

Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, *15*(5), 447–452.

Friedman, O., & Petrashek, A. R. (2009). Children do not follow the rule "ignorance means getting it wrong." *Journal of Experimental Child Psychology*, *102*(1), 114–121.

Hamlin, K. J., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785.

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*(3), 567–582.

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*, 72–80.

Huang, Z., Hu, Q., & Shao, Y. (2019). Understanding others' knowledge certainty from inference and information-seeking behaviors in children. *Developmental Psychology*, *55*(7), 1372.

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015a). The goggles experiment: Can chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour*, *105*, 211–221.

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015b). Chimpanzees strategically manipulate what others can see. *Animal Cognition*, *18*(5), 1069–1076.

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, *87*(3), 179–208.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149.

Marticorena, D. C., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science*, *14*(6), 1406–1416.

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375–382.

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40–48.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Mossler, D. G., Marvin, R. S., & Greenberg, M. T. (1976). Conceptual perspective taking in 2-to 6-year-old children. *Developmental Psychology*, *12*(1), 85.
Nordmeyer, A. E., & Frank, M. C. (2014). The role of context in young children's comprehension of negation. *Journal of Memory and Language*, *77*, 25–39.
O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*(2), 659–677.
Reuter, T., Feiman, R., & Snedeker, J. (2018). Getting to no: Pragmatic and semantic factors in two- and three-year-olds' understanding of negation. *Child Development*, *89*(4), e364–e381.
Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind & Language*, *11*(4), 388–414.
Sullivan, K., & Winner, E. (1991). When 3-year-olds understand ignorance, false belief and representational change. *British Journal of Developmental Psychology*, *9*(1), 159–171.
Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.

Intersubjectivity and social learning: Representation of beliefs enables the accumulation of cultural knowledge

Carles Salazar

Department of Art and Social History (Anthropology Program), Faculty of Arts, University of Lleida, Plaça Víctor Siurana, 1, E25003 Lleida, Spain.
Carles.salazar@udl.cat;
http://www.hahs.udl.cat/ca/personal-academic/pagina-1/carles-salazar-carrasco/

doi:10.1017/S0140525X20001648, e168

Abstract

I accept the main thesis of the article according to which representation of knowledge is more basic than representation of belief. But I question the authors' contention that humans' unique capacity to represent belief does not underwrite the capacity for the accumulation of cultural knowledge.

The authors make a very good point in demonstrating the fundamental nature of knowledge representation in humans. It has older evolutionary origin than that of belief representation, and that explains why nonhuman primates can do the first but fail to do the second. But is it not a contradiction to argue, on the one hand, that knowledge representation, in so far as it can be seen as a basic cognitive competence, is not distinctive of the human species and, on the other, that what we normally see as the most distinctive characteristic of the human species, which is the capacity to accumulate cultural knowledge, originates in that very same competence? If this is so, one could legitimately wonder why cumulative cultural knowledge is not much more widespread among nonhuman primates than what seems to be the case (Tennie, Call, & Tomasello, 2009; Whiten, 2017).

The authors only mention the accumulation of cultural knowledge at the end of the paper, in section 6.2.1, and they do not elaborate the reasons why they confidently state that "Although the ability to represent others' beliefs may indeed turn out to be unique to humans and critically important for some purposes, it does not seem to underwrite humans' capacity for the accumulation of cultural knowledge." However, this is undoubtedly a key question for all the sciences of human behavior. A priori, one could plainly state that knowledge representation, rather than belief representation, is instrumental to the accumulation of cultural knowledge for the very simple reason that it is "knowledge" what we accumulate, not "beliefs." Does that mean that understanding beliefs is irrelevant in the process of social learning that leads to the accumulation of cultural knowledge?

Belief representation, the authors concede, is relevant for predicting other people's behavior, but it is knowledge, and not belief, "that allows us to represent others as reliable guides to the actual world" (6.1). This is undoubtedly true in a rather obvious sense; but it can also be misleading, for it glosses over the process of social learning as it takes place in all known human societies and that enables any apprentice to acquire knowledge from his or her teacher (Sterelny, 2012). Let me illustrate this with a very simple example. If I want to know how a computer works, I may ask a computer scientist about it. Quite obviously, I am interested in the computer scientist's knowledge about computers, not about her beliefs. But the point I wish to make is that I shall only have access to that knowledge if I am able to understand her beliefs (Salazar, 2018, pp. 37–62).

There is ample evidence that the process of social learning among humans is not simply learning from others, but it is normally conducted within some form of pre-existing social bond (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981; Henrich, 2015; Kline, 2015; Nielsen, 2008; Zuidema, 2002). More specifically, when social learning entails the transmission of socially shared forms of knowledge, what we normally define as "culture," social learning can only take place when some culturally significant form of social relationship links teacher and apprentice. For the majority of human societies, these social relationships are normally kinship relationships and, more specifically, family relationships, for it is from those that the first and most elementary parts of one's cultural knowledge are to be acquired (Demps, Zorondo-Rodríguez, García, & Reyes-García, 2012; Hewlett & Cavalli-Sforza, 1986; McElreath & Strimling, 2008). This basic nucleus of kinship relations will later be supplemented by other kinds of relationships in different ways. WEIRD (western, educated, industrialized, rich, and democratic) societies are somewhat unique in the sense that they have reduced the social relationship between teacher and learner to the (relatively) impersonal bond created in institutional schooling. However, even when there is some form of selectivity (Bentley & O'Brien, 2011), cultural knowledge is very rarely transmitted between anonymous individuals (cf. Osiurak & Reynaud, 2020).

But why should that be the case? One might be tempted to argue that those networks of social relationships provide a sort of external framework within which "real" knowledge can circulate, but they do not really affect the nature of that knowledge in any substantial way and, crucially, do not transform it into "mere beliefs." Let me show why this cannot be a valid assumption by going back to the simple example of the teacher – computer scientist. The knowledge I am likely to obtain from her will certainly be a partial knowledge about how the computer actually works – otherwise, I would become a computer scientist myself. But, given my ignorance about computers, there is no way I can have access to that knowledge if I have not previously understood what she *believes* to be the case about the computer and, specifically, if I do not trust her (Csibra & Gergely, 2006; see Hewlett, Fouts, & Boyette, 2011). In other words, before getting knowledge from any teacher, I have to believe in that teacher and share her intentionality, so that my knowledge becomes a "dialogic cognitive representation" (Tomasello, Carpenter, Call, Behne, & Mol, 2005). In order to acquire the objective knowledge about the world that will enable me to make use of my computer,

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

I have to understand what goes on in the mind of the computer scientist that is teaching me, that is, understand her beliefs so I can end up thinking "through her mind" (Veissière, Constant, Ramstead, Friston, & Kirmayer, 2020). This is what identifies cultural learning as a specific form of social learning (Tomasello, Kruger, & Ratner, 1993).

To conclude, from an objective point of view, the accumulation of cultural knowledge does effectively entail knowledge representation. But subjectively, that accumulation is only possible through belief representation. Culture is knowledge acquired from a subject, not from the world, hence only humans' capacity to understand other minds as "subjects in the world" permits its assimilation. And it is by understanding another subject's beliefs that I can assimilate her knowledge and, consequently, I can add up her knowledge to mine, that is, accumulate.

Conflict of interest

None.

References

Bentley, R. A., & O'Brien, M. J. (2011). The selectivity of social learning and the tempo of cultural evolution. *Journal of Evolutionary Psychology*, *9*(2), 125–141.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution. A quantitative approach*. Princeton University Press.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition. The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development. Attention and performance* (pp. 249–274). Oxford University Press.

Demps, K., Zorondo-Rodríguez, F., García, C., & Reyes-García, V. (2012). Social learning across the life cycle: Cultural knowledge acquisition for honey collection among the Jenu Kuruba, India. *Evolution and Human Behavior*, *33*, 460–470.

Henrich, J. (2015). *The secret of our success*. Princeton University Press.

Hewlett, B. S., & Cavalli-Sforza, L. L. (1986). Cultural transmission among Aka pigmies. *American Anthropologist*, *88*, 922–934.

Hewlett, B. S., Fouts, H. N., & Boyette, A. H. (2011). Social learning among Congo Basin hunter-gatherers. *Phil. Trans. R. Soc. B*, *366*, 1169–1178.

Kline, M. A. (2015) How to learn about teaching: An evolutionary framework for the study of teaching behaviour in humans and other animals. *Behavioral and Brian Sciences*, *e31*, 1–71. doi:10.1017/S0140525X14000090.

McElreath, R., & Strimling, P. (2008). When natural selection favors imitation of parents. *Current Anthropology*, *49*(2), 307–315.

Nielsen, M. (2008). The social motivation for social learning. *Behavioral and Brain Sciences*, *31*, 33. doi:10.1017/S0140525X0700324X.

Osiurak, F., & Reynaud, E. (2020). The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences*, *43*, e156, 1–66. doi:10.1017/S0140525X19003236.

Salazar, C. (2014). Understanding belief: Some qualitative evidence. *Journal of Empirical Theology*, *2*(27), 199–213.

Salazar, C. (2018). *Explaining human diversity. Cultures, minds, evolution*. Routledge.

Sterelny, K. (2012). *The evolved apprentice. How evolution made humans unique*. MIT Press.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B*, *364*, 2405–2415.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Mol, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–735.

Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495–511. doi: 10.1017/S0140525X0003123X.

Veissière, S. P. L., Constant, A., Ramstead, M. J. D., Friston, K. J., & Kirmayer, L. J. (2020). Thinking through other minds: A variational approach to cognition and culture. *Behavioral and Brain Sciences*, *43*, e90, 1–75. doi: 10.1017/S0140525X19001213.

Whiten, A. (2017). A comparative and evolutionary analysis of the cultural cognition of humans and other apes. *The Spanish Journal of Psychology*, *19*, e98, 1–19.

Zuidema, W. (2002). The importance of social learning in the evolution of cooperation and communication. *Behavioral and Brain Sciences*, *25*, 283–284.

Teleology first: Goals before knowledge and belief

Tobias Schlicht[a], Johannes L. Brandl[b], Frank Esken[c], Hans-Johann Glock[d,e], Albert Newen[a], Josef Perner[f], Franziska Poprawe[d], Eva Schmidt[g], Anna Strasser[h], and Julia Wolf[a]

[a]Institute of Philosophy II, Ruhr-Universität Bochum, GA3/29, 44780 Bochum, Germany; [b]Department of Philosophy KGW, University of Salzburg, A-5020 Salzburg, Austria; [c]University of Europe for Applied Sciences, 58636 Iserlohn, Germany; [d]Philosophisches Seminar, Universität Zürich, CH-8044 Zürich, Switzerland; [e]Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, CH-8044 Zurich, Switzerland; [f]Department of Psychology, Centre for Cognitive Neuroscience, 5020 Salzburg, Austria; [g]Department of Philosophy and Political Science, TU Dortmund, 44227 Dortmund, Germany and [h]Independent Researcher, 10245 Berlin, Germany
tobias.schlicht@rub.de, www.rub.de/philosophy/situatedcognition
johannes.brandl@sbg.ac.at, http://www.johannesbrandl.com
frank.esken@ue-germany.com
glock@philos.uzh.ch, https://www.isle.uzh.ch/en.html
albert.newen@rub.de
josef.perner@sbg.ac.at, https://ccns.sbg.ac.at/people/perner/
franziska.poprawe@philos.uzh.ch, https://www.philosophie.uzh.ch/de/seminar/people/research/theory_glock/poprawe.html
eva.schmidt@tu-dortmund.de
annakatharinastrasser@gmail.com
https://www.denkwerkstatt.berlin/ANNA-STRASSER/
Julia.Wolf-n8i@ruhr-uni-bochum.de

doi:10.1017/S0140525X20001533, e169

Abstract

Comparing knowledge with belief can go wrong in two dimensions: If the authors employ a wider notion of knowledge, then they do not compare like with like because they assume a narrow notion of belief. If they employ only a narrow notion of knowledge, then their claim is not supported by the evidence. Finally, we sketch a superior teleological view.

We are sympathetic to the author's focus on "understanding others' minds in relation to the actual world" (p. 2), because it recognizes that references to worldly facts rather than mental states are primary in explaining actions. However, the empirical evidence cited does not support their main claim "knowledge before belief." Our criticism can be put in terms of a dilemma: (1) If the authors do *not* employ the same notion of knowledge throughout the paper, then they do not compare like with like. Although the intended focus of the argument is on declarative *knowledge-that*, some passages (sects 4.1 and 6.1) employ a wider notion including less demanding *knowledge-how*. Consequently, the authors should also consider the possibility of a wider, non-propositional form of belief. (2) If they *do* only employ *knowledge-that*, then their claim is not supported by the evidence, which is better explained by more basic means. (3) A superior *teleological* account grounds action explanation in an appreciation of an agent pursuing a goal.

(1) The authors specify four features "essential to knowledge" (p. 4), but do not elaborate on their operative conception of

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

belief. Because they assess ascription of belief via understanding *false* belief, they assume a demanding notion of belief throughout, without considering a wider notion setting weaker constraints on belief. The features of knowledge mentioned are not characteristic of *knowledge-how*. *Knowledge-how* to swim can be mastered better or worse but does not amount to knowing a fact. By analogy, a wider notion of belief could be characterized as a minimally structured informational state that can be systematically connected to motivational states (see Newen & Starzak, 2020). Comparing like with like suggests a parallelism between knowledge and belief. If we ascribe propositional *knowledge-that*, we employ a concept that has "merely" *believing-that* as a fallback option, such that both unfold as a package from more basic roots.

(2) Although the authors formulate their central claim concerning the primacy of propositional *knowledge-that*, most of the evidence mentioned in section 4 can be explained in terms of perceptual access and/or *knowledge-how*. Understanding others in terms of perceptual access is simpler because it is immediately situation-based. To understand that Eve reaches for something because she *sees* it, I can rely on her line of sight. This explains the study by Krachun, Carpenter, Call, and Tomasello (2009) without recourse to an attribution of knowledge. The chimpanzees learn to look for the food where their human competitors look for it because they recognize that their competitors could observe both where the food was placed and the switch of the containers. Similarly, the experiment by Luo and Johnson (2009) modulates an agent's perceptual access to information and does not warrant attribution of *knowledge-that*. Furthermore, the studies by Behne, Liszkowski, Carpenter, and Tomasello (2012) and Kovács, Tauzin, Téglás, Gergely, and Csibra (2014) do not license the conclusion that infants attribute to others *knowledge-that* which they lack. Infants may simply recognize that the other agent "encountered" an object (Butterfill & Apperly, 2013) or that she *knows how* to acquire, interact with, or refer to an object. This suggests that there is a basic capacity for representing *knowledge-how* without yet involving *knowledge-that*. In the first instance, we learn from others *how to do things*, not facts. This is borne out by the evidence provided in section 6.2: Nonhuman primates gain *knowledge-how* to forage or to solve problems; and in order to learn children turn to *competent* adults, that is, to adults with *know-how*.

(3) Regarding the question of how we explain others' actions, there is an even simpler alternative, ignored by the authors. "Teleology" postulates a basic way of understanding simple actions. Developmental evidence suggests that young infants ground their expectations about people's actions in perceived objects and facts and that adults continue to use such explanatory strategies wherever it proves sufficient (Gergely & Csibra, 2003; Perner & Roessler, 2010). Facts are reasons for action (Alvarez, 2010). The structure of the information that the interpreter attributes to the agent is not yet differentiated into attributing states of believing and states of knowing (see Perner & Esken, 2015; Perner & Roessler, 2012; Roessler & Perner, 2013). By understanding an agent's action as a means of pursuing a goal, the interpreter recognizes facts motivating the agent's actions without representing her mental states. Only at around the age of four do children learn to appreciate that other agents can relate to facts differently from their own perspective, manifest, for example, in the attribution of false beliefs (Wimmer & Perner, 1983).

If the authors employ a wide notion of knowledge including *knowledge-how*, then teleology is more informative by bringing in facts and goal-directed actions. If the authors employ a demanding *knowledge-that* account, then teleology offers more parsimonious explanations of the relevant data, without relying on a demanding understanding of either belief *or* knowledge.

The teleological approach also has more specific advantages over a knowledge-first account:

(a) It captures both *informational* and *motivational* aspects of action understanding (Glock & Schmidt, 2019) and allows for further developmental steps toward a more sophisticated understanding of others.

(b) It offers a realistic conception of how humans eventually arrived at ideas such as "belief" and "knowledge." Animals can act for reasons – on account of facts – without understanding that others have such reasons too (Glock, 2019). Similarly, one need not represent an agent's mental state in order to understand that she acts on account of facts.

(c) It allows action predictions in cases where appealing to mental states (knowledge or belief) lacks warrant. Where will my colleague be at 4pm today? No idea what goes through her mind, but at 4pm is our faculty meeting. Because this fact provides reasons for her to attend, I predict she'll be there.

(d) Teleology provides an explanation for the "reality error," which is more economical than that of the knowledge-first approach. Children anticipate that someone will look for the object in its real location because he or she has objective reasons to do so.

In short, instead of "knowledge first," we suggest "teleology first," that is, sensitivity to facts is fundamental.

Financial support. This commentary grew out of a joint research project: "The structure and development of understanding actions and reasons." TS and AN are grateful to the German Science Foundation (DFG, SCHL 588/3-1; NE 576/14-1); HG, ES, and FP would like to thank the Swiss National Science Foundation (SNF, #5100019E_177630) and HG also the NCCR "Evolving Language" (Swiss National Foundation #51NF40_180888); and JP and FE thank the Austrian Science Fund (FWF, I 3518-G24). JW's contribution is part of her work in the Research Training Group "Situated Cognition," funded by the German Science Foundation (DFG, GRK 2185/1).

References

Alvarez, M. (2010). *Kinds of reasons*. Oxford University Press.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, *30*, 359–375.

Butterfill, S. A., & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind and Language*, *28*(5), 606–637.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*, 287–292.

Glock, H. (2019). Agency, intelligence and reasons in animals. *Philosophy (London, England)*, *94*, 645–671.

Glock, H., & Schmidt, E. (2019). Objectivism and causalism about reasons for action, with E. Schmidt. In G. Schumann (Ed.), *Explanation in action theory and historiography: Causal and teleological approaches. Routledge studies in contemporary philosophy* (pp. 124–145). Routledge.

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy*, *19*(6), 543–557.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149.
Newen, A., & Starzak, T. (2020). How to ascribe beliefs to animals. *Mind and Language*, 1–19. https://doi.org/10.1111/mila.12302.
Perner, J., & Esken, F. (2015). Evolution of human cooperation in homo heidelbergensis: Teleology versus mentalism. In P. Barrouillet (Ed.), *Recent advances in cognitive-developmental theory. Special issue, developmental review* (Vol. *38*, pp. 69–88). Elsevier.
Perner, J., & Roessler, J. (2010). Teleology and causal reasoning in children's theory of mind. In J. Aguilar & A. Buckareff (Eds.), *Causing human action: New perspectives on the causal theory of action* (pp. 199–228). MIT Press.
Perner, J., & Roessler, J. (2012). From infant's to children's appreciation of belief. *Trends in Cognitive Sciences*, *16*(10), 519–525.
Roessler, J., & Perner, J. (2013). Teleology: Belief as perspective. In S. Baron-Cohen, H. Tager-Flusberg & M. Lombardo (Eds.), *UOM-3: Understanding other minds* (3rd ed., Chapter 3, pp. 35–50). Oxford University Press.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5.

There's more to consider than knowledge and belief

David M. Sobel

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912, USA.
Dave_Sobel@brown.edu

doi:10.1017/S0140525X20000989, e170

Abstract

Phillips et al. present a number of arguments for the premise that knowledge is more basic than belief. Although their arguments are coherent and sound, they do not directly address numerous cases in which belief appears to be a developmental precursor to knowledge. I describe several examples, not necessarily as a direct challenge, but rather to better understand their framework.

Phillips, Buckwalter, Cushman, Friedman, Martin, Turri, Santos, and Knobe review findings from comparative, developmental, and cognitive psychology on the general theme that knowledge is more basic in our conception than belief. Overall, I find their general argument convincing, particularly as it relates to developmental progressions, evidenced by, for example, the theory of mind scales (Wellman & Liu, 2004). I would like to see them give a fairer consideration of their alternative "View 1" – the hypothesis that belief attribution is more basic than knowledge attribution (at least regarding developmental progressions). I will present two cases not described in their article to see whether they pose challenges for the framework they propose, and discuss the broader implications of these challenges.

1. Knowledge and belief in pretense

Some have argued that children's ability to pretend demonstrates early representational competence and that children scaffold their representational understanding of pretense to help them make explicit judgments about others' false beliefs (e.g., Leslie, 1987). Others, however, have suggested that although young children engage in pretense, doing so posits only the same representational capacities as moving one's body; pretense is "acting-as-if" (Lillard, 1993a; Nichols & Stich, 2003; Perner, 1991). Support for this perspective comes from variants of the "Moe the troll" paradigm: Children are shown a troll doll (Moe), who is hopping up and down like a kangaroo. Because there are no kangaroos in the land of the trolls, Moe doesn't know what one is, and has never seen one before. Four-year-olds – who pass explicit false-belief measures – will erroneously say that Moe is pretending to be a kangaroo (Lillard, 1993b). Here is a case of judgments about others' (false) beliefs being made developmentally earlier than judgments of others knowledge, particularly as they relate to pretending.

Even if one rejects the acting-as-if hypothesis (Friedman & Leslie, 2007) or suggests that the Moe findings reflect children's broader causal reasoning (Sobel, 2009), there is a large body of research that suggests children generally understand false belief prior to their understanding the relation between knowledge and other mental states (reviewed in Lillard, 2001; Stich & Tarzia, 2015). Notable for the present argument is Perner, Baker, and Hutton's (1994) concept of *prelief*: On this view, knowledge is not more basic than belief. Rather, pretense and belief are an undifferentiated concept when pretend play emerges. They become differentiated with success on false-belief measures, but prelief itself seems more basic than a concept of knowledge.

2. Knowledge and belief in selective learning

Phillips et al. point out that children selectively learn from others, based on their evaluations of their epistemic competence. They conclude, however, that selective learning relies on "representations of knowledge rather than belief in determining from whom to learn." It is not clear how they come to this conclusion. Classic measures of selective learning (e.g., Koenig, Clément, & Harris, 2004; Koenig and Harris, 2005) introduced preschoolers to two informants. One labeled familiar objects accurately. The other labeled the same objects inaccurately. Researchers used three different measures: (1) *Explicit Judgments* questions about the informants – whether one informant was either a good or bad labeler. (2) *Endorse* questions in which they were shown novel objects that were given different novel labels by each informant (e.g., one labeled it a dax, the other a wug); children were asked whether they thought the object was a dax or a wug. (3) *Ask* questions in which children were asked from whom they wanted to learn labels of novel objects.

These questions potentially ask different things about the knowledge and belief states of the informants. Ask questions assess what children believe about the two informants' knowledge (i.e., given the demonstrations of epistemic competence you've observed, from whom would you want to learn?). Explicit Judgment questions assess a valence judgment about the informants' knowledge. Endorse questions, in contrast, assess what children believe the label of the object really is (presumably, what children believe the informants believe the label to be). Success on these questions – children's ability to use information from informants selectively – has distinct developmental trajectories. Meta-analyses now suggest that at the youngest ages tested, children perform well on Endorse questions, whereas performance on Ask questions and Explicit Judgment questions develops significantly during the preschool years (Sobel & Finiasz, 2020; Tong, Wang, & Danovitch, 2020). Children's selective learning about the belief states of others seems to be present quite

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

early. Selective inferences about facets of others' knowledge seem to have prolonged developmental trajectories.

3. Do I believe everything I know?

I've focused on instances of belief being more basic than knowledge. There are potentially others. Older preschoolers fail on certain measures of true belief, even when they pass measures of false belief (see Hedger & Fabricius, 2011). False-belief contrastive utterance ("I thought it was an X, but it was a Y") emerge before children pass false-belief measures (see Bartsch & Wellman, 1995), and lead to the possibility that children's understanding of knowledge itself changes (the "connectionist" construal described on pp. 54–55). Linguistic analysis of adults' usage of the words *know* and *think* to children suggest that the data necessary to recognize that *know* is factive is sparse (e.g., Dudley, Rowe, Hacquard, & Lidz, 2017).

Therefore, although I suspect the view that Phillips et al. advocate has merit, I'd like to see them more carefully consider the alternative account, particularly from a developmental perspective. The findings I've mentioned here require integration into the framework they have set up, as they shed doubt on the hypothesis that all aspects of knowledge are understood by children earlier than belief, and in some cases, suggest that children's conceptualization of knowledge and belief changes over development. Further integrating their arguments with other developmental findings would make their "call to action" to study the role of knowledge in theory of mind development more compelling.

Financial support. The author was funded by NSF grants 1661068, 1917639, and 2033368 during writing of this commentary.

Conflict of interest. None.

References

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford University Press.

Dudley, R., Rowe, M., Hacquard, V., & Lidz, J. (2017). Discovering the factivity of "know." *Semantics and Linguistic Theory*, *27*, 600–619.

Friedman, O., & Leslie, A. M. (2007). The conceptual underpinnings of pretense: Pretending is not "behaving-as-if." *Cognition*, *105*(1), 103–124.

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Review of Philosophy and Psychology*, *2*(3), 429.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*(4), 412.

Lillard, A. S. (1993a). Pretend play skills and the child's theory of mind. *Child Development*, *64*(2), 348–371.

Lillard, A. S. (1993b). Young children's conceptualization of pretense: Action or mental representational state? *Child Development*, *64*(2), 372–386.

Lillard, A. (2001). Pretend play as twin earth: A social-cognitive analysis. *Developmental Review*, *21*(4), 495–531.

Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Clarendon Press/Oxford University Press.

Perner, J. (1991). *Understanding the representational mind*. MIT Press.

Perner, J., Baker, S., & Hutton, D. (1994). Prelief: The conceptual origins of belief and pretence. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 261–286). Erlbaum.

Sobel, D. M. (2009). Enabling conditions and children's understanding of pretense. *Cognition*, *113*(2), 177–188.

Sobel, D. M., & Finiasz, Z. (2020). How children learn from others: An analysis of selective word learning. *Child Development*, *91*(6), e1134–e1161.

Stich, S., & Tarzia, J. (2015). The pretense debate. *Cognition*, *143*, 1–12.

Tong, Y., Wang, F., & Danovitch, J. (2020). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science*, *23*(2), e12895.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.

Two distinct concepts of knowledge

Christina Starmans

Department of Psychology, University of Toronto, Toronto, Ontario M5S 3G3, Canada.
christina.starmans@utoronto.ca

doi:10.1017/S0140525X20001892, e171

Abstract

The central claim in the target article is that representations of knowledge are more basic than representations of beliefs. However, the authors are blending together two distinct concepts of knowledge: "awareness" and "propositional knowledge." Distinguishing these two concepts of knowledge clarifies how the developmental and comparative data fit within the philosophical literature.

In this provocative and important paper, two starkly opposing theories of the relationship between knowledge and belief are pitted against one another. View 1 represents the way knowledge has long been understood in epistemology: Representations of knowledge depend on prior representations of belief, and therefore the attribution of belief is more psychologically "basic" than the attribution of knowledge. View 2 is the surprising idea that representations of knowledge are more basic than representations of belief.

The authors argue that a wealth of evidence from developmental, comparative, and social psychology, as well as from experimental philosophy, supports view 2 and refutes view 1. However, I suggest that both of these views are actually correct, and that the evidence supporting view 2 doesn't undermine the longstanding view of knowledge captured by view 1. This is because the word "knowledge" in English corresponds to two concepts (indeed, other languages use multiple words for different types of knowing that capture some of this polysemy); the concept of knowledge referred to in view 1 is a different concept of knowledge than the one referred to in view 2.

View 1 concerns the concept of propositional knowledge that has long been a topic of interest in epistemology. To know that P requires a justified belief that P, and it also requires that P is true. (It also likely requires additional factors; see Gettier, 1963; Starmans & Friedman, 2012.) The appropriate contrast class for this concept of knowledge is belief. That is, a question about knowledge-belief asks: "Does she know that P, or does she merely believe that P?" Hence, when considering propositional knowledge, it's impossible to attribute knowledge without first attributing belief, because having a belief is just one component of having knowledge.

View 2 concerns knowledge in the sense of "awareness." To have knowledge in this sense only requires not being ignorant. A question about awareness asks: "Does she know that P, or is she ignorant of P?" When considering whether someone is ignorant, the issue isn't the presence or absence of a belief; it is whether or not there is awareness.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

The research reviewed in the target article reveals that infants, young children, and nonhuman primates attribute knowledge in this latter sense. And, as carefully illustrated by Phillips et al., this does not require attributing belief. Infants successfully predict how others will behave when aware of an object's location (usually, although not always, through visually observing it). However, when an observer's perceptual access to the location of an object has been disrupted, although adults would expect that they still have a true belief, infants now act as if they are ignorant, and have no expectation as to their behavior. Similarly, chimpanzees have demonstrated an understanding of awareness and ignorance, but there is no clear evidence that they are able to represent the beliefs of others, whether true beliefs or false beliefs. Furthermore, there are circumstances in which adults more readily attribute knowledge than belief (e.g., Myers-Schulz & Schwitzgebel, 2013; Phillips et al., 2018), suggesting that this notion of knowledge persists through development. All of this supports view 2.

However, there is another sense of knowledge that adults are concerned with. Perhaps the most famous quest for knowledge of this sort was that of Descartes, who asked how he could know whether the things he believed to be true were actually true. Descartes had plenty of beliefs that he was very confident about, such that he was sitting in his dressing gown by the fire, holding writing papers in his hands. But he worried that he might be dreaming, or insane, or deceived by a malignant demon. He wanted to know whether these beliefs rose to the level of knowledge, and ultimately concluded that there was one thing he did know: "I knew that I was a substance whose whole essence or nature is simply to think, and which does not require any place or depend on any material thing, in order to exist" (Descartes, 1980/1637, p. 18). Here, Descartes is clearly referring to propositional knowledge, not simply awareness.

Phillips et al. might argue that this notion of knowledge is special to philosophers; knowledge in the "awareness" sense is the only concept that exists in everyday use. But this clearly isn't the case. Ordinary people ask all the time whether someone knows something or just believes it. This might mean asking whether someone is certain of their belief, whether someone's belief is true, or whether they have the right kind of evidence for their belief. In all these cases, however, we are first attributing a belief to someone, and then asking whether this belief rises to the level of knowledge. We might also withhold attribution of knowledge even in cases where there is awareness. Someone in the desert, half mad with thirst, might see a (real) lake in front of him – might be fully *aware* of it – but worry that it's a hallucination. "I believe there is water over there, but I just don't know for sure!" he might cry out to himself.

Properly distinguishing these two concepts of knowledge reveals two things. First, representations of knowledge in the sense of awareness are more basic than representations of belief. Second, representations of belief are more basic than representations of propositional knowledge. View 2 is correct, but so is view 1.

Conflict of interest. None.

References

Descartes, R. (1980/1637). *Discourse on method and meditations on first philosophy*. Donald A. Cress trans. Hackett Publishing Co.

Gettier, E. (1963). Is justified true belief knowledge?. *Analysis*, *23*(6), 121–123.

Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing that P without believing that P. *Noûs*, *47*(2), 371–384.

Phillips, J., Strickland, B., Dungan, J., Armary, P., Knobe, J., & Cushman, F. (2018). Evidence for evaluations of knowledge prior to belief. *Proceedings of the Fortieth Annual Conference of the Cognitive Science Society*.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, *124* (3), 272–283.

Are knowledge- and belief-reasoning automatic, and is this the right question?

Andrew D. R. Surtees[a] and Andrew R. Todd[b]

[a]School of Psychology, University of Birmingham and Birmingham Children's Hospital, Edgbaston, Birmingham B15 2TT, UK and [b]Department of Psychology, University of California, Davis, Davis, CA 95616, USA.
A.Surtees@Bham.ac.uk; atodd@ucdavis.edu
https://www.birmingham.ac.uk/staff/profiles/psychology/surtees-andrew.aspx, https://psychology.ucdavis.edu/people/atodd

doi:10.1017/S0140525X20001880, e172

Abstract

Phillips et al. conclude that current evidence supports knowledge-, but not belief-reasoning as being automatic. We suggest four reasons why this is an oversimplified answer to a question that might not have a clear-cut answer: (1) knowledge and beliefs can be incompletely equated to perceptual states, (2) sensitivity to mental states does not necessitate representation, (3) automaticity is not a single categorical feature, and (4) how we represent others' minds is dependent on social context.

The target article makes an important theoretical contribution. Comparing knowledge and belief representation provides a compelling account that will shape future research significantly. Phillips et al. rightly note that research on the automaticity of adult belief and knowledge representation is contentious. We remain sceptical that knowledge versus belief representation will ever neatly classify as automatic or not. We consider four aspects of Phillips et al.'s reasoning to illustrate this point.

1. Visual perspectives are not pure analogies to knowledge and belief representations

Phillips et al. invoke the distinction between level-1 and level-2 perspectives taking to distinguish knowledge versus belief representation. The strongest evidence supporting adults' automatic knowledge representation comes from a level-1 perspective taking task (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010), whereas level-2 tasks are used to support the non-automaticity of belief reasoning (Surtees, Butterfill, & Apperly, 2012; Surtees, Apperly, & Samson, 2016a). Both sides of the analogy between visual perspectives and knowledge versus belief are problematic, however. It is not necessarily the case that someone's current level-1 perspective is consistent with their knowledge state. Although *seeing is knowing* is a reasonable heuristic (Moll & Tomasello, 2007), *not seeing is not knowing* is unlikely to be. We rarely beep our horn as our neighbours reverse towards their houses without looking, because we know they *know* it is there. Regarding level-2 perspective taking, representing how

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

someone sees something does not necessarily equate to representing their belief about the object. You can look at a number 6 from one angle, although we see it as a number 9 from another, without us holding differing *beliefs* about the object. By equating knowledge and belief with level-1 and level-2 perspective taking, respectively, Phillips et al. may be over-simplifying knowledge and over-complicating belief.

2. Mental-state sensitivity may not necessitate mental-state representation

Phillips et al. conclude evidence of automaticity based on studies documenting interference from another's perspective on judgements of one's own perspective (Kovács, Téglás, & Endress, 2010; Samson et al., 2010). We question whether the incidental sensitivity to others' mental states revealed by these altercentric-interference effects necessitates their being *represented*. Several accounts leave open the possibility that other mechanisms underlie such mental-state sensitivity. As Phillips et al. note, submentalizing accounts propose lower-level, domain-general explanations of altercentric interference (Heyes, 2014). Two-systems accounts, in contrast, posit that altercentric interference reflects domain-specific *registration* of "belief-like states" (Apperly & Butterfill, 2009). Taking cues from both of these accounts, process-dissociation accounts hold that altercentric interference is not a process-pure index of mental-state registration; rather, such effects can be decomposed into at least two component processes: calculation of the agent's perspective and detection of one's own perspective (Todd, Cameron, & Simpson, 2017, in press; Todd, Simpson, & Cameron, 2019), with the latter process likely reflecting something more domain-general (Payne, 2005). Phillips et al. reason that knowledge representation is more basic based on evidence from tasks finding altercentric interference, but this could be evidence of arbitrary correlation between "knowledge" and stimulus features, coupling to more basic mental-state-like states, or poorer discrimination from self-knowledge.

3. Automaticity is not categorical

Phillips et al. aim to categorize knowledge and belief representation as automatic or not. Such categorization is likely an over-simplification of how cognitive systems operate. Automaticity is not a single feature, but rather a set of conceptually separable features that often do not co-occur (Melnikoff & Bargh, 2018; Moors & De Houwer, 2006). Thus, specifying *in what way(s)* belief and knowledge representation are automatic is crucial. The automaticity features receiving most empirical attention are goal-independence and efficiency. We agree with Phillips et al. that current evidence supports level-2 altercentric interference and process-dissociation estimates of agent-perspective calculation as consistently not-automatic. On a level-2 task, when participants only ever considered their own perspective, we showed that altercentric interference (Surtees et al., 2016a) and agent-perspective calculation (Todd et al., in press) were absent, suggesting level-2 perspective taking is goal-dependent. Using the same task, Todd et al. (2019) found that time pressure also impaired agent-perspective calculation, suggesting it is relatively inefficient. Evidence for level-1 automaticity is more equivocal. Some studies suggest level-1 altercentric interference emerges regardless of participant task goals (Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Surtees et al., 2016a); others do not (Ferguson, Apperly, & Cane, 2017; Todd et al., in press). Some studies suggest level-1 altercentric interference and agent-perspective calculation are efficient, in that they were unimpaired by time pressure (Todd et al., 2017, 2019) or a concurrent resource-consuming task (Qureshi, Apperly, & Samson, 2010); another found the opposite (Qureshi & Monk, 2018). Although Phillips et al. acknowledge empirical uncertainty, our view is that different mental-state representations are not necessarily categorized fully as automatic or non-automatic. The impact of context on automaticity further supports this contention.

4. Mental-state representation may differ in observational and interactive contexts

Processes for interaction are different when directly engaging with another person, as opposed to passively observing them (Schilbach et al., 2013). Phillips et al. largely focus on observational, "third-person" approaches to adults' mental-state representation. It is not self-evident that the more "basic" form of mental-state reasoning in interactive and observational scenarios will be equivalent. Here, level-2 perspective taking provides an example of where they are not. Above we highlighted that level-2 perspective taking in an observational setting seems to be goal-dependent (Surtees et al., 2016a; Todd et al., in press). In a closely matched interactive context, however, participants do suffer interference from a partner's perspective, even if they are never asked to report it, suggesting it is goal-independent (Elekes, Varga, & Király, 2016; Surtees, Samson, & Apperly, 2016b). One possibility is that representations of different mental states are differentially cued by different aspects of social interaction. In this case, perhaps low-level stimulus features cue level-1 perspective taking, whereas "real" interaction or a specific goal may be required to cue level-2 perspective taking. One way to interpret this is that level-1 perspective taking is more basic, but another is to see level-2 perspective taking as linked to social interaction in a more fundamental way. Classifying knowledge or belief representation as more basic may obfuscate subtle variation with context.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953.

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 454.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, *41*, 93–103.

Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology*, *70*(8), 1646–1660.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131–143.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293.

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*, *43*(2), 309.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297.

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, *89*(4), 488.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.
Qureshi, A. W., & Monk, R. L. (2018). Executive function underlies both perspective selection and calculation in level-1 visual perspective taking. *Psychonomic Bulletin & Review*, *25*(4), 1526–1534.
Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255.
Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and Brain Sciences*, *36* (4), 393–414.
Surtees, A., Apperly, I., & Samson, D. (2016a). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52.
Surtees, A., Samson, D., & Apperly, I. (2016b). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105.
Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, *30*(1), 75–86.
Todd, A. R., Cameron, C. D., & Simpson, A. J. (2017). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, *159*, 97–101.
Todd, A. R., Cameron, C. D., & Simpson, A. J. (in press). The goal-dependence of level-1 and level-2 visual perspective calculation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
Todd, A. R., Simpson, A. J., & Cameron, C. D. (2019). Time pressure disrupts level-2, but not level-1, visual perspective calculation: A process-dissociation analysis. *Cognition*, *189*, 41–54.

Knowledge-by-acquaintance before propositional knowledge/belief

Michael Tomasello

Department of Psychology and Neuroscience, Duke University, Durham, NC 27708, USA.
Michael.tomasello@duke.edu

doi:10.1017/S0140525X20001387, e173

Abstract

More basic than the authors' distinction between knowing and believing is a distinction between knowledge-by-acquaintance (I know John Smith) and propositional knowledge/belief (I know/believe *that* John Smith lives in Durham). This distinction provides a better account of both the comparative and developmental data.

The authors distinguish between two cognitive processes: believing and knowing. But this is not the most basic or productive distinction to be made in the current context. The most basic distinction, in my view, is between two kinds of knowing, indicated in many of the world's languages by two different verbs: in German by *kennen* and *wissen*, in French by *connaître* and *savoir*, and in Spanish by *conocer* and *saber*. The first member of each pair is normally glossed as something like "be acquainted with," as in "I know (am acquainted with) the restaurant to which you are referring." The second member of each pair is normally glossed in propositional terms, as in "I know *that* the restaurant to which you are referring is on 9th Street." The opposite of knowledge-by-acquaintance is ignorance: "I do not know (am ignorant of) the restaurant to which you are referring." The opposite of propositional knowledge is error: "The restaurant to which you are referring is not on 9th Street."

My claim is that this distinction between knowledge-by-acquaintance and propositional knowledge is most basic. So what about believing? Believing is inherently propositional. One does not believe a restaurant; rather, one believes *that* a particular restaurant is closed. Believing involves an agent representing some state of the world, and either she or some outside observer has some reason to doubt that this representation is accurate – as in all propositional judgments. Thus, you *know* that the restaurant to which I am referring is on 9th Street, and you *believe* (i.e., either I or you are not sure that this is the case) that its ZIP Code is 27709. The difference is that I agree with your first judgment (or else you are sure), but I disagree (or you are not sure) about your second judgment. The key point is that believing and knowing in the propositional sense both involve cognitive representations of states of affairs coupled with judgments about the accuracy or inaccuracy of the representation as compared to the real situation (i.e., one can have different "attitudes" to the propositional content). Such comparison between representation and reality is not involved in knowledge-by-acquaintance. Even if we establish that I am acquainted with some state of affairs – I am acquainted with the fact that the restaurant is on 9th Street – this does not involve an attitude or judgment about whether or not my representation of that state of affairs is accurate.

My empirical claim is that what comes first in both phylogeny and ontogeny is an understanding that agents are acquainted with things (mainly by perception). Thus, chimpanzees understand that a competitor sees, and so is acquainted with, the location of a piece of food, and they even understand that a competitor has seen food being hidden and thereby has become acquainted with its location (Hare, Call, Agnetta, & Tomasello, 2000; Hare, Call, & Tomasello, 2001). (NB: the apes have propositional knowledge, but they do not attribute it to others.) One might choose to gloss this as an understanding that the competitor knows *that* the food is in a certain location; but this would not fit with the empirical data. Studies specifically designed to distinguish between knowledge-by-acquaintance and propositional knowledge in great apes (e.g., Kaminski, Call, & Tomasello, 2008) show that when chimpanzees see a competitor witnessing the hiding of food, they understand that she is acquainted with the food's location; when they witness her not witnessing the hiding of food, they attribute to her ignorance. But the key finding is this: When chimpanzees witness a competitor being blocked from witnessing the moving of food from one location to another, they do not attribute to her an incorrect representation of the food's location based on her now outdated witnessing, but only, again, ignorance of the food's location. They can understand that their competitor is acquainted with the food's location, but they do not compare her representation of the situation either to their own or to any "objective" situation, and so there is no question of accuracy or potential error. They attribute to their competitor knowledge-by-acquaintance (or ignorance), not propositional knowledge, belief, or error.

The same analysis applies to looking studies with human infants (and apes). As argued by Tomasello (2018), in the classic looking-time studies of infants' understanding of false belief (e.g., Onishi & Baillargeon, 2005), it may be that infants simply understand that the agent is acquainted with (has registered) the location of the object based on where she last saw it. But, to do this, infants do not need to relate or coordinate their understanding of the agent's representation with any other representation, neither their own nor some objective representation. To be surprised or to predict that the agent is searching for the object somewhere other than where she saw it disappear does not require a judgment of whether her knowledge is accurate or inaccurate; the infant's or ape's own

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

knowledge of the location of the object is irrelevant and not attended to (and the same analysis applies to anticipatory looking studies, such as that of Krupenye, Kano, Call, Hirata, & Tomasello, 2016).

I thus agree that an understanding that others know things is both phylogenetically and ontogenetically primary, but only if we are talking about knowledge-by-acquaintance involving simple representations. Understanding representations as propositional entails, in addition, an understanding that they either match or mismatch with the objective situation as represented by the one making the judgment. Propositional knowledge and beliefs thus involve a comparison and/or coordination of different representations of one and the same situation, which presupposes both a prior understanding of something more primitive like knowledge-by-acquaintance (based on simple perception and representation) and, in addition, an ability to compare and/or coordinate potentially different representations (at an executive level).

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*, 771–785.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behavior*, *61*, 139–151.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*, 224–234.

Krupenye, C., Kano, F., Call, J., Hirata, S., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*, 110–114.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, *115*, 8491–8498.

The evolution of knowledge during the Cambrian explosion

Walter Veit

School of History and Philosophy of Science, The University of Sydney, Camperdown, NSW 2006, Australia.
wrwveit@gmail.com; https://walterveit.com/

doi:10.1017/S0140525X20000771, e174

Abstract

Phillips et al. make a compelling case for a reversal in the current paradigm in "other minds" research by considering the representation of other people's knowledge more basic than the attribution of belief. Unfortunately, they only discuss primates. In this commentary, I argue that the representation of others' knowledge is an evolutionary ancient trait, first appearing during the Cambrian explosion.

In their target article "Knowledge before belief" by Phillips et al., we are presented with a radical reversal of the current paradigm in "other minds" research. Breaking with a long tradition that sought to understand the minds of other humans (and animals) by focusing on the attribution of beliefs, the authors argue that decades of empirical research in the cognitive sciences have undermined or at least begun to call into question the assumption that the attribution of knowledge rests on a more basic or fundamental capacity to attribute beliefs. For historical, methodological, and philosophical reasons, however, other minds research has long been held back from even considering this option in the conceptual space.

One way to formulate the underlying problem is to ask which way of representing the minds of others came first – the representation of knowledge or of beliefs? By *first* here I mean something stronger than just during the course of human development, that is, first in the sense of being evolutionarily more ancient. Unfortunately, Phillips et al. have little to say about the evolutionary history of these traits and, perhaps more worryingly, seem to conflate evolution and development, discussing both under the banner of whether the representation of other people's knowledge occurs first or later in human infancy. But the order of appearance of these traits in human development may not be the same as in evolutionary history. Ontogeny does not have to track phylogeny. By not paying heed to this fallacy, however, evolutionarily problematic conclusions straightforwardly follow. If one assumes, for instance, that representation of beliefs come developmentally prior in humans, one will only grant representation of others' knowledge to those animals that are also able to also represent the others' beliefs. But, as Phillips et al. themselves recognize, the latter ability may turn out to be unique to humans. This would then lead to the phylogenetically untenable conclusion that humans are the only creatures on this planet able to represent the mental states of others.

Naturally, there are multiple ways out of this dilemma – and the most attractive one will certainly be to outright reject the notion that the ability to represent others' beliefs comes first. Phillips et al. accumulate supporting evidence from nonhuman primate species to make the case that the human ability to represent beliefs is phylogenetically recent (Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011; Martin & Santos, 2014, 2016), but I think that they could have dived much deeper into our evolutionary history to support their case.

Approximately 541 million years, in fact, for this is the beginning of the Cambrian explosion when most animal body plans first appeared (Maloof et al., 2010). The ability to track other's knowledge is, I shall argue, an evolutionary ancient trait appearing roughly at the beginning of the Cambrian. What is notable in the early Cambrian is an increase in body size and the emergence of various sensory modalities to track one's environment. But more sophisticated ways of sensing one's surrounding naturally led ways of sensing others – to react. This emergence of a richer kind of agency gave rise to arms races between predators and prey (Bengtson, 2002) and the evolution of centralized nervous systems (Wray, 2015) to coordinate action and perception. It is during this special period that some philosophers and scientists locate the origins of subjectivity and subjective experience (Godfrey-Smith, 2017; Ginsburg & Jablonka, 2019). In research on animal consciousness, there is a temptation to look for human indicators – signs of conscious experience that are perhaps unique to human life. But such approaches give rise to views that draw firm boundaries between us and other animals (Veit & Huebner, 2020), a problem that is similarly present in research on the origins of other minds' representation. To switch from the rich intentional belief attribution to the perhaps computationally simpler knowledge attribution may reveal a picture in which the latter is evolutionarily truly ancient. Daniel Dennett's *intentional stance*

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

programme has long emphasized that the ability to attribute beliefs should not be conceived as the sudden emergence of a new sophisticated faculty in human, but one that is similarly present in other animals (Dennett, 1987; Veit et al., 2020). Now, we may have to recognize that it should have been the attribution of knowledge to others that deserves our attention here.

An important observation made by Godfrey-Smith (2016) is that there is a transition somewhere in the Cambrian after which "the mind evolved in response to *other* minds" (p. 63). This transition should be understood as the evolution of representing other minds' knowledge. An important question for both predator and prey becomes: *Have I been seen?* The existence of eyes appears to function as a shorthand for many animals to make just this inference – when eyes meet, one infers knowledge of ones' location to the subject at the other end of this exchange. Burrowing, ink release, and flight are useful attempts to break this link. Many predators avoid the eye contact of their prey at all cost. Knowledge and ignorance of one's surroundings can make all the difference to survival. The evolution of eye-spots on butterflies is one spectacular invention to make potential predators think that they are seen, thus avoiding conflict. Behaviourists may appeal to simpler explanations, but in this case, knowledge attribution may not be such a complex affair. To see others in one's environment as subjects is bound to give one an edge over others in an ecology of interaction. But to treat others as subjects entails the attribution of knowledge.

The picture I have offered here is a speculative one – one that ties the explosion of diversity during the Cambrian to the recognition of other minds' knowledge. Nevertheless, it offers additional support to the main conclusion in Phillips et al. Focusing on human representations of other minds might have biased us against a much more basic approach to other minds research. The attribution of knowledge to other minds may be an evolutionarily much more ancient trait than the attribution of belief, an idea that I will follow up elsewhere (Veit 2021).

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. None.

References

Bengtson, S. (2002). Origins and early evolution of predation. *Paleontological Society Papers*, *8*, 289–318. doi: 10.1017/S1089332600001133.

Dennett, D. C. (1987). *The intentional stance*. MIT Press.

Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul: Learning and the origins of consciousness*. MIT Press.

Godfrey-Smith, P. (2016). Animal evolution and the origins of experience. In D. Livingstone Smith (Ed.), *How biology shapes philosophy: New foundations for naturalism* (pp. 51–71). Cambridge University Press.

Godfrey-Smith, P. (2017). The subject as cause and effect of evolution. *Interface focus*, *7* (5), 20170022. https://doi.org/10.1098/rsfs.2017.0022.

Maloof, A. C., Porter, S. M., Moore, J. L., Dudas, F. O., Bowring, S. A., Higgins, J. A., … Eddy, M. P. (2010). The earliest Cambrian record of animals and ocean geochemical change. *Geological Society of America Bulletin*, *122*(11–12), 1731–1774.

Marticorena, D. C. W., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science*, *14*(6), 1406–1416. doi: 10.1111/j.1467-7687.2011.01085.x.

Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition*, *130*(3), 300–308.

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375–382.

Veit, W. (2021). *Health, agency, and the evolution of consciousness*. Ph.D. thesis, University of Sydney. Manuscript in preparation.

Veit, W., Dewhurst, J., Dołega, K., Jones, M., Stanley, S., Frankish, K., & Dennett, D. C. (2020). The rationale of rationalization. *Behavioral and Brain Sciences*, *43*, e53. https://doi.org/10.31234/osf.io/b5xkt.

Veit, W., & Huebner, B. (2020). Drawing the boundaries of animal sentience. *Animal Sentience*, *29*(13), 1–4. https://animalstudiesrepository.org/animsent/vol5/iss29/13/.

Wray, G. A. (2015). Molecular clocks and the early evolution of metazoan nervous systems. *Philosophical Transactions of the Royal Society B*, *370*, 20150046. doi: 10.1098/ rstb.2015.0046.

Why is knowledge faster than (true) belief?

Evan Westra

Department of Philosophy, York University, Toronto, ON M3J 1P3, Canada.
ewestra@yorku.ca; https://sites.google.com/site/ewestraphilosophy

doi:10.1017/S0140525X20001399, e175

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. Although 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 explain 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. Because 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 of *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

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

demands executive functions such as working memory and inhibitory control (Fizke, Barthel, Peters, & Rakoczy, 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: Although 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 solely because of 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, Boyer, Weimer, & Carroll, 2010; Horschler, Santos, & MacLean, 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 although 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. Although 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 from 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, 2021).

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 types of cases 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, because 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 (cf. Rakoczy, Bergfeld, Schwarz, & Fizke, 2015). In these cases, the processing demands (and hence, the speed) of belief attribution and knowledge attribution should be equally demanding.

Financial support. This study was supported by Social Sciences and Humanities Research Council of Canada Postdoctoral Fellowship 756-2018-0012.

Conflict of interest. None.

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.

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*, 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. (2021). Mindreading in conversation. *Cognition*, *210*, 104618. https://doi.org/10.1016/j.cognition.2021.104618.

Theory of mind in context: Mental-state representations for social evaluation

Brandon M. Woo[a], Enda Tan[b], and J. Kiley Hamlin[b]

[a]Department of Psychology, Harvard University, Cambridge, MA 02138, USA; and [b]Department of Psychology, University of British Columbia, Vancouver, BC, Canada V6T 1Z4.
bmwoo@g.harvard.edu
enda.tan@psych.ubc.ca
kiley.hamlin@psych.ubc.ca
https://bmwoo.github.io/
https://cic.psych.ubc.ca/

doi:10.1017/S0140525X20001818, e176

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

Abstract

Whereas Phillips and colleagues argue that knowledge representations are more basic than belief representations, we argue that an accurate analysis of what is fundamental to theory of mind may depend crucially on the context in which mental-state reasoning occurs. Specifically, we call for increased study of the developmental trajectory of mental-state reasoning within socially evaluative contexts.

To support their argument that knowledge representations are more basic than belief representations, Phillips et al. draw on evidence suggestive that knowledge representations emerge earlier and more robustly in infancy than belief representations. They propose that knowledge representations, unlike belief representations, are developmentally privileged and "fundamental because they allow us to learn from others about [the true state of] the world." Here, we argue that learning the true state of the world is one, but not the only, function of theory of mind; therefore, it may be premature to conclude that knowledge is more fundamental than beliefs. Specifically, we argue for an increased focus on the role of mental-state representations in contexts involving the evaluation of potential social partners (i.e., socially evaluative contexts).

Humans represent mental states not only to learn from others about the true state of the world, but also to learn about the character of potential social partners within it. Because humans must cooperate with each other to survive (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012), we must be able to accurately assess potential social partners and determine whether they might cooperate with us in the future. Our theory of mind is crucial in this process, enabling us to distinguish, for instance, between an individual who intentionally poisoned someone's coffee and one who did so under the false belief that the poison was sugar (Young, Cushman, Hauser, & Saxe, 2007). Which individual would make a better social partner? Our ability to represent others' mental states, including not only what they know, but also what they believe, is critical for evaluating others' actions and readily informs partner choice decisions (see Martin & Cushman, 2015).

Despite early developing motivations to form and maintain social relationships (see Raz & Saxe, 2020), the vast majority of studies on theory of mind development have not examined mental-state representations in socially evaluative contexts. Rather, studies on infant theory of mind have almost exclusively assessed infants' expectations of a single, neutral agent who seeks to find an object. Phillips and colleagues include these studies as part of their evidence that knowledge representations are more basic than belief representations.

Without grounding studies of infants' mental-state representations in contexts of social evaluation, however, it may be premature to form such conclusions. A large body of research has demonstrated that the context of a task may matter for false-belief reasoning as well as cognitive reasoning more broadly. For example, adults' cognitive reasoning is enhanced when tasks are framed as being about social contracts versus in non-social terms (Cosmides & Tooby, 1992). Similarly, a number of studies suggest that young children (who typically struggle in verbal tasks of false-belief understanding) may be better able to answer questions about false beliefs when agents act antisocially (Chandler, Fritz, & Hala, 1989; Tsoi, Hamlin, Waytz, Baron, & Young, 2020; Wellman, Cross, & Watson, 2001). Here, we explore the possibility that when infants engage in social evaluation, they show earlier capacities for mental-state reasoning than previously assumed.

Take, for instance, studies assessing 3-month-old infants' understanding of others' goals. Past research has found that 3-month-olds do not readily represent the goals of agents' object-directed actions (Sommerville, Woodward, & Needham, 2005). And yet 3-month-olds do appear to negatively evaluate agents who hinder others' goal pursuit: They selectively avoid looking at agents who steal a protagonist's ball (Hamlin & Wynn, 2011) and who prevent a protagonist's attempts to climb up a hill (Hamlin, Wynn, & Bloom, 2010). These findings point to the possibility that 3-month-old infants may be more capable of representing others' (even unfulfilled) goals in socially evaluative versus non-evaluative contexts. Given these findings within the domain of goal understanding, do socially evaluative contexts facilitate infants mental-state representations more broadly?

Indeed, a growing number of studies have provided evidence that infants' social evaluations incorporate others' intentions and knowledge states by late in the first year (Hamlin, 2013; Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; Woo, Steckler, Le, & Hamlin, 2017), and recent research suggests that they may even incorporate others' false beliefs by 15 months. In Woo and Spelke (2020), 15-month-olds evaluated agents not on the basis of the objective consequences of their actions (whether they caused a protagonist to obtain a desired versus an undesired toy), but instead on the basis of whether or not the agents *believed* their actions would be helpful. That is, infants preferred an agent who directed a protagonist to a location where the agent had last seen a toy that they knew the protagonist desired (i.e., where the agent falsely believed the desired toy to be), over an agent who inadvertently directed the protagonist to the desired toy's actual location (i.e., where the agent falsely believed the desired toy was not). These findings replicated in two distinct testing contexts (in-person and online), and provide the first evidence that infants can reason about false beliefs in socially evaluative contexts.

Although these results clearly require replication by independent researchers, they stand in contrast to the mixed evidence that infants represent false beliefs in non-evaluative contexts (Poulin-Dubois et al., 2018). In light of research suggestive that the development of infants' goal understanding may also differ across socially evaluative and non-evaluative contexts, we call on future studies to systematically test whether the development of mental-state representation differs across contexts. Given the importance of mental states, both factive and non-factive, for accurate social evaluation, infants may be more sensitive to what others believe earlier in development in socially evaluative contexts than in the non-evaluative contexts of traditional false-belief tasks.

In sum, we argue that the study of theory of mind must consider the context in which mental-state reasoning occurs. Although both knowledge and belief have major consequences for social evaluation, the vast majority of studies on infants' theory of mind to date have not examined infants' mental-state representations in socially evaluative contexts, but instead in a comparatively inconsequential object search paradigm. By examining the development of belief representations in a wider range of contexts, we can better determine which mental states are fundamental to theory of mind.

Financial support. BW was supported by a Social Sciences and Humanities Research Council Doctoral Fellowship under award 752-2020-0474.

Conflict of interest. None.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

References

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development*, *60*, 1263–1277.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford University Press.

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), 451–474.

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30–39.

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–929.

Martin, J. W., & Cushman, F. (2015). To punish or to leave: Distinct cognitive processes underlie partner control and partner choice behaviors. *PLOS ONE*, *10*(4), e0125193.

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., … Perner, J. (2018). Do infants understand false beliefs? We don't know yet–a commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, *48*, 302–315.

Raz, G., & Saxe, R. (2020). Learning in infancy is active, endogenously motivated, and depends on the prefrontal cortices. *Annual Review of Developmental Psychology*, *2*.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*, 673–692.

Tsoi, L., Hamlin, J. K., Waytz, A., Baron, A. S., & Young, L. (2020) *False belief understanding for negative versus positive interactions in children and adults*. https://moralitylab.bc.edu/wp-content/uploads/2020/09/Tsoi_MeanNiceAnne_children_adults.pdf.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Woo, B. M., & Spelke, E. (2020). Infants' social evaluations depend on the intentions of agents who act on false beliefs. PsyArXiv. https://doi.org/10.31234/osf.io/eczgp.

Woo, B. M., Steckler, C. M., Le, D. T., & Hamlin, J. K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition*, *168*, 154–163.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240.

Authors' Response

Actual knowledge

Jonathan Phillips[a], Wesley Buckwalter[b], Fiery Cushman[c], Ori Friedman[d], Alia Martin[e], John Turri[f], Laurie Santos[g], and Joshua Knobe[h]

[a]Program in Cognitive Science, Department of Psychological and Brain Sciences and Department of Philosophy, Dartmouth College, Hanover, NH 03755, USA; [b]Department of Philosophy, Institute for Philosophy and Public Policy, George Mason University, Fairfax, VA 22030, USA; [c]Department of Psychology, Harvard University, Cambridge, MA 02138, USA; [d]Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1, Canada; [e]School of Psychology, Victoria University of Wellington, Wellington 6012, New Zealand; [f]Philosophy Department and Cognitive Science Program, University of Waterloo, Waterloo, ON N2L 3G1, Canada; [g]Department of Psychology, Yale University, New Haven, CT 06520, USA and [h]Program in Cognitive Science, Department of Philosophy, Yale University, New Haven, CT 06520, USA.

jonathan.s.phillips@dartmouth.edu; https://phillab.host.dartmouth.edu/
wesleybuckwalter@gmail.com; https://wesleybuckwalter.org/
cushman@fas.harvard.edu; http://cushmanlab.fas.harvard.edu/
friedman@uwaterloo.ca; https://sites.google.com/view/uwaterloocclab
alia.martin@vuw.ac.nz; https://vuwbabylab.com/
john.turri@gmail.com; https://john.turri.org/
laurie.santos@yale.edu; https://caplab.yale.edu/
joshua.knobe@yale.edu; https://campuspress.yale.edu/joshuaknobe/

doi:10.1017/S0140525X21000911, e177

Abstract

This response argues that when you represent others as knowing something, you represent their mind as being related to the actual world. This feature of knowledge explains the limits of knowledge attribution, how knowledge differs from belief, and why knowledge underwrites learning from others. We hope this vision for how knowledge works spurs a new era in theory of mind research.

R1. Introduction

Since the publication of Premack & Woodruff's (1978) article "Does the chimpanzee have a theory of mind?" in this journal, researchers have taken the point of theory of mind to be determining the content of others' thoughts – what exactly it is they think or want. And it has become an accepted truism that this capacity is essentially for predicting and explaining others' behavior (the literature, *passim*). The central and basically only point we are going to make in this response is that this way of understanding theory of mind has gotten it all backward.

We think that the capacity for *theory of mind*, in its most basic form, is not primarily concerned with the content of others' thoughts which sometimes happens to reflect the actual world; it's primarily concerned with the content of the actual world, which other minds happen to reflect. The signature features of this basic capacity – that it is factive, that it requires more than justified true belief, that it allows you to represent others as knowing *more* than you, and that it is not modality specific – suggest that the capacity did not evolve specifically for predicting and explaining others' behavior. After all, the ability to represent someone as knowing where they hid the cookies from you, for example, isn't particularly useful for predicting where they'll go to get the cookies. It's true that there are cases in which this basic capacity will be useful for determining the content of others' thoughts or predicting and explaining their behavior, but our point is that the signature features of knowledge make it clear that this basic form of theory of mind did not evolve for this purpose, in particular. Instead, we've argued, the signature features of this basic capacity suggest it evolved to help you interact with and learn from others, precisely because it allows you to keep track of what they understand about the actual world.

So despite what a few philosophers said in their commentaries to Premack and Woodruff's article now more than 40 years ago, the core cases in theory of mind research should never have been ones involving false beliefs. Much more revealing are ones in which someone else is better informed than you, rather than worse. Therefore, if you're in the business of reading *Behavioral and Brain Sciences* in search of paradigms that separate one's own representation of the world from others' understanding of it, these seem like good ones that might reveal something about our basic capacity for theory of mind. Perhaps, it's finally time to let go of false beliefs and focus instead on the way things actually are.

A number of the commentaries to our target article objected to this way of understanding theory of mind. Some objected to the

idea that theory of mind representations concern the actual world rather than the contents of others' minds (sect. R2). Others argued that the signature limitations on this basic form of theory of mind simply boil down to limits on what kind of *content* one can or does attribute to others' minds (sects. R3 and R4). Still others objected that theory of mind really is for predicting and explaining others' behavior, as opposed to coordinating on the actual world or learning about it from others (sect. R5).

By and large, we think these commentaries are onto something. In fact, we think that if you can just accept that knowledge is concerned with the actual world, you'll get to have your cake and eat it too. There are limitations on what kind of content you represent others as knowing, but those boil down to limitations on the kind of content you think the actual world involves. And sometimes this basic capacity for theory of mind does let you figure out the content of others' thoughts, but those are just the cases in which you also happen to know whatever it is someone else knows. If you *do* happen to know the answers to a test question (say you wrote the exam), you'll know a great deal about what is going on in the minds of the students who also know the answers. In such cases, you'll even be pretty decent at predicting what answers they'll give and explaining why they gave those answers. But, in all the other cases, where you don't know what it is that someone else knows, the exact same capacity will still be incredibly useful – this time for learning about the actual world itself, which is one of the things that the capacity for knowledge representation is particularly well-designed for.

Other commentaries saw the merit in our basic vision for theory of mind and came up with a surprising number of elegant suggestions for improvement. We did our best to take these to heart, and we'll point out the many places where they've made the central argument clearer or more convincing, or have expanded this basic vision toward new horizons (sects. R6 and R7). So much for the introduction, and now on to the details of how to actually understand knowledge.

R2. Knowledge and the actual world, or, the truth

As we were at pains to point out in the target article, knowledge, unlike belief, is factive. While there are different ways to spell out the factivity condition on knowledge (see the commentary by **Nagel**), what is not controversial is that it ensures that you cannot represent others as knowing anything you take the actual world to preclude. Thus, there is some uncontroversial correspondence between your own understanding of the actual world and your attribution of knowledge to others. But what does this correspondence amount to?

A number of the commentaries propose that this basic form of theory of mind essentially involves your own understanding of the world (among others, see the commentaries by **Durdevic & Krupenye**, **Tomasello**, and **Westra**). On this view, when you represent others as knowing something, you represent their mind as being related to the actual world as you understand it. You yourself take the actual world to be some way, and representing someone as knowing something involves representing them as having the right kind of relation to that part of the world. The things that they know, then, are things involved in your own understanding of the world. As emphasized in the commentary by Tomasello, a form of theory of mind with this structure might be as simple as tracking whether other agents are acquainted with physical parts of the world, such that they know where the ball is, or know the woman who knitted mittens for your niece.

One alternative proposal takes a step back from the actual world and proposes that this basic form of theory of mind essentially involves monitoring whether others' understanding conflicts with your own (see, e.g., the commentary by **Deschrijver**). The actual world might happen to inform one's own understanding, but does not play a direct role, because knowledge attribution takes place at the level of tracking the relationship between two minds: your own mind and another's.

A third approach, which we might think of as the farthest from actuality, argues that this basic form of theory of mind, if it is a genuine form of theory of mind, must be meta-representational in the same way that belief is (Leslie, 1987). That is, it must essentially involve you representing someone else's independent representation of the world, and accordingly is not essentially concerned with the actual world as you understand it. With varying degrees of commitment, versions of this approach are discussed by **Dudley & Kovács**; **Gordon**; **Kampis & Csibra**, and **Binmore**.

So why do we think knowledge involves one's own understanding of the actual world? One reason is that if you don't think this, you don't have a natural way of accounting for why knowledge but not belief is factive. If attributing knowledge (but not belief) involves understanding others' minds in relation to the actual world, then the factivity of knowledge comes for free. Obviously, you cannot understand someone's mind as being related to some part of the actual world when you think the actual world contains no such part. If knowledge attribution is instead meta-representational in the same way belief is, then some extra explanation must be given for why this representation just happens to be limited by precisely the bounds of your own understanding of the actual world. It's not that you couldn't give such an account, but we can't see why you would want to. There's a much simpler explanation on offer.

This way of explaining the factivity of knowledge is importantly different from the version proposed by **Nagel**. Nagel proposes that the factivity constraint on knowledge is *modal*: Knowledge *necessarily* only binds agents to truths. The natural way of understanding Nagel's suggestion is that when people represent someone as knowing something, they understand the person's mental states as having this property (necessarily being true). From our perspective, the trouble with this approach is that the data suggest that knowledge attribution is unlikely to involve representing any such modal property. Nonhuman primates, for instance, seem to have a remarkable capacity to attribute knowledge (see sect. 4.1), but we'd be pretty shocked if they have the capacity to represent anything as being necessarily true. And if they don't have the capacity to represent anything as necessarily being true, then a fortiori they don't have the capacity to represent others' mental states as necessarily being true. So we can be pretty sure that if nonhuman primates do attribute knowledge, knowledge does not involve reasoning about which truths hold across possible worlds. And there's evidence that human adults aren't all that different (Turri, 2018). In fact, humans seem relatively happy to attribute knowledge in exactly the cases that philosophers designed to illustrate that knowledge cannot be attributed when such modal properties are violated (see, e.g., Colaço, Buckwalter, Stich, and Machery, 2014, on fake barn intuitions and Turri, 2016a on reliabilism). Such attributions are to be expected if knowledge concerns the actual world, not merely possible ones. So the first and perhaps most obvious reason to think knowledge involves the actual world is that this gives you a simple explanation of both why and how knowledge is factive.[1]

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

A second reason to find this approach promising is that it also explains why the capacity for knowledge representation is more basic than the capacity for belief representation (see the commentary by **Westra** for a related line of reasoning). If knowledge attribution involves understanding others' minds in relation to the world, then one can maintain a single representation of the world, parts of which others also know about. Belief, unlike knowledge, cannot involve one's own understanding of the world in the same way because beliefs can be false. Thus, belief, unlike knowledge, requires an independent representation of the world – the world merely as understood by another – which must be maintained separately from one's own understanding, and thus can be false.

Once again, another feature of knowledge – its comparative basicness – falls naturally out of our way of understanding knowledge attribution, while other approaches leave this feature unexplained. Consider, for example, the suggestion by **Lassiter** that the basic theory of mind capacity we presented may be better explained by representations of *true* belief rather than knowledge, or the suggestion from **Sobel** that the evidence may be better explained with the notion of *prelief* (i.e., representations that are understood to not be real, but also are not understood to be false, as in pretense). What remains perplexing is why attributions of true belief or prelief, which require the same resources as genuine belief representation, would show all of the signature features of a more basic cognitive capacity: emerging early in phylogeny, ontogeny, processing time, and may be processed automatically and persist in the face of other cognitive impairments.

Of course, one could go on to give some further explanation of why some additional difficulty emerges in cases of false beliefs in particular, for example, **Deschrijver** proposes a difficulty with conflict monitoring. However, such explanations face the challenging task of carving apart the cases that are genuinely difficult from cases that seem similarly complex but are not as difficult. Consider, for example, a recent piece of empirical evidence that demonstrates a nuanced capacity for attributing knowledge in monkeys (Horschler, Santos, & MacLean, 2019). In this experiment, monkeys watched an experimenter who saw a piece of fruit move into one of two containers in a display in front of them. A screen then blocked the view of the experimenter and one of two things occurred. In half of the conditions, the fruit itself briefly moved out of the container and then back inside. In the other half of the conditions, the fruit remained where it was, but the container briefly moved off of the fruit and then back on it. In both cases, all objects had returned to the position where the experimenter had last seen them, and using looking time, researchers investigated whether the monkeys expected the experimenter to reach for the fruit where it was last seen. What Horschler and colleagues found was that when the box moved, monkeys continued to expect the experimenter to reach for the fruit where they had last seen it. However, when the fruit moved instead, monkeys no longer expected the experimenter to reach for the fruit where they had last seen it.

If your account proposes a difficulty with conflict monitoring (**Deschrijver**) or that one can only represent true beliefs (**Lassiter**), these results are worryingly hard to explain. Such accounts focus on representations that are independent from the actual world. Thus, when the experimenter doesn't know about them, things that happen in the actual world shouldn't change what the experimenter believes. Accordingly, the most natural prediction for such accounts is that in *both* conditions the experimenter will be represented as having a belief about the location of the fruit, and this belief will happen to be true – it will match the monkey's own ideas about the location of the fruit. So, in both conditions, monkeys should expect the experimenter to reach for the fruit where it actually is. But, of course, monkeys don't do that. Instead Horschler and colleagues found that monkeys only expect the experimenter to reach for the fruit when it was the container, rather than the fruit, that moved.

The difference between the conditions is easy enough to explain, however, if monkeys represent knowledge rather than belief. Knowledge requires more than having a justified true belief (sect. R2; Gettier, 1963). And so when the fruit moves but the experimenter doesn't see it (but then happens to return to the original location), the experimenter might end up with a true belief about the location of the fruit by coincidence, but they do not share the monkey's understanding of the location of the fruit. By contrast, when only the container moves, this should *not* affect the experimenter's knowledge of the fruit, and monkeys should continue expecting the experimenter to act in accordance with this knowledge. This is exactly what they do.

This is just one of a growing number of studies that demonstrate clear failures to represent others' true beliefs while simultaneously demonstrating clear success in representing their knowledge (see Horschler, Santos, & MacLean, 2021; Krachun, Carpenter, Call, & Tomasello, 2009). The key difference is that knowledge tasks can be passed by simply keep track of whether the agent understands the relevant part of the actual world, whereas the true belief tasks require you to construct a separate representation of the world as the agent understands it, which just happens to align with your own understanding, and thus is true.

Therefore, in short, if you can accept that knowledge concerns the actual world, you get a surprisingly simple explanation for why knowledge is basic, why it is factive, and how it differs from belief.

R3. But what do we know anyway?

Instead of locating the difference between knowledge and belief in the role of one's own understanding of the world, a number of commentaries argued that the essential difference between them concerns the kind of content they allow you to attribute. After all, as **Tomasello** and **Starmans** point out, human languages typically encode an intriguing difference between knowledge and belief. In English (as in many other languages), one can know ways of doing things and know the smell of summer rain, but one cannot *believe* ways of doing things or *believe* the smell of summer rain. Tomasello and Starmans argue that the content of belief attributions seems to be propositional, while the content of knowledge attributions can be both actual things in the world and abstract matters of fact.

Following their line of thought further, it wouldn't be surprising if there were different mechanisms for understanding, on the one hand, the kind of acquaintance other agents have to physical parts of the world and, on the other hand, their acquaintance with things like abstract propositional truths. Moreover, it is plausible enough that the mechanisms for figuring out what physical parts of the world another is acquainted with may be simpler than the mechanisms for figuring out which propositions another is acquainted with. And as **Tomasello** and **Starmans** point out, much of the evidence for basic knowledge ascriptions in nonhuman primates and human infants suggest that these populations represent others as knowing about physical objects or having certain skills. So perhaps all of this points to a key distinction in kinds of knowledge, with a basic form of knowledge attribution

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

that amounts to little more than knowledge-by-acquaintance or know-how and differs sharply from belief attribution, and a separate more complex form of propositional knowledge ascription that is not more basic than belief ascription but is rather quite similar to it. On this view, the difference in basicness we illustrated in the target article is a matter of the basicness of the content attributed (propositional vs. non-propositional), and not truly a matter of the basicness of the attitude itself (knowledge vs. belief). This all seems quite convincing.

The trouble is that there actually seems to be a simpler explanation for why there isn't great evidence that nonhuman primates and human infants represent others as having knowledge of abstract propositions. Namely, there isn't great evidence that nonhuman primates and human infants represent abstract propositions, in general. If knowledge attributions essentially involve your own understanding of the world, then the kind of content one can represent others as knowing will depend on what kind of content your own representation of the actual world involves.

Moreover, the similarity between propositional and non-propositional content shouldn't be hard to see here. If you do not understand the actual world to involve any extraterrestrial aliens, you could not represent anyone as knowing them ("knowledge-by-acquaintance"). And if you do not think there are ways of turning water into gold bullion, you can't represent anyone as knowing how to do that (knowledge-how). And in just the same way, if you do not understand the actual world to involve abstract propositions, like "2 + 7 = 10" then you certainly will not be able to represent others as knowing this sort of thing either ("propositional knowledge").

Nonhuman primates (and perhaps very young human infants) may not have the capacity to represent propositions, and thus their knowledge representations will necessarily be restricted to simpler forms of content, whether knowledge-by-acquaintance (**Tomasello**) or even just visual perspective (**Asaba, Chuey, & Gweon** [**Asaba et al.**]). And if this is right, then such knowledge representations are also likely to be guided by specific attention to cues such as eye gaze or direct perception (**Dudley & Kovács**; **Grossmann & Dela Cruz**; **Kano & Call**). However, for human adults who clearly can and do represent the world in something closer to propositional terms, the same capacity may be used to represent others as having knowledge of abstract truths. For example, as emphasized beautifully by **Mikhail**, human adults represent others as having moral and legal knowledge. Although unquestionably abstract, these rules make up part of our understanding of the world, and given that, we have no trouble representing others as sharing our understanding of them. Note that in the latter kinds of cases, we agree with **Westra** that there is reason to think the content of knowledge is propositional and with **Farina & Lavazza** who argue that knowledge is content-involving and representational.

Importantly although, even for unambiguously propositional content, attributing knowledge seems easier than attributing belief. One completely uncontroversial piece of evidence is that young children succeed at attributing propositional knowledge (e.g., "Sally does not know her marble is in the basket") before they succeed in attributing similarly propositional beliefs (e.g., "Sally believes her marble is in the box"). Similarly, adults are faster to correctly attribute or deny knowledge claims than they are to correctly attribute or deny corresponding belief claims, even when the term used for knowledge is explicitly propositional, for example, "savoir" in French, which only takes propositional complements (Phillips, Knobe, Strickland, Armary, & Cushman, 2018). A third piece of evidence comes from the commentary by **Bricker**, who used electroencephalogram (EEG) to show that propositional knowledge representation elicited a weaker P3b amplitude than belief representation (Bricker, 2020). Thus, even when knowledge attributions unambiguously involve propositional content, they continue to show signs of emerging earlier, being simpler, and requiring less processing than matched belief attributions.

So it turns out the surprisingly simple solution is that the mechanism for representing knowledge is just the same across all of these different kinds of cases – you are just figuring out what parts of the world someone else understands – and seeming differences in the complexity of knowledge attributions across species or development arise simply from the complexity of representing different parts of the actual world. (See **Rosenbaum, Halilova, & Pathman** for related commentary on the difference in complexity between episodic and semantic content in knowledge attribution.)

The upshot of our view is that knowledge attributions won't be limited to any particular type of content (propositional, knowledge-how, etc.). Knowledge attributions can be as rich as your own understanding of the world. It is for this reason that we suspect that the capacity for knowledge attribution we provided evidence for in the target article will be not be fully captured by approaches that place limits on the content of basic theory of mind attributions, for example, reducing it to representations of visual perspective (**Asaba et al.**), uninterrupted perceptual access (**Dudley & Kovács**), skill (**Carpendale & Lewis**), goals (**Schlicht, Brandl, Esken, Glock, Newen, Perner, Poprawe, Schmidt, Strasser, & Wolf** [**Schlicht et al.**]), episodic experience (**Kampis & Csibra**), or knowledge-by-acquaintance (**Tomasello**). While each of these commentaries does an excellent job of pointing to specific aspects of knowledge we can attribute to others, it would be quite surprising on each of these views if knowledge attribution just happened to work in much the same way in all the other cases as well. That is, each of these cases shares the signature features of knowledge attribution (sect. 2 of the target article). We don't think this is surprising though. Each of these cases involve various aspects of the actual world, and representing others as knowing that part of the world will work similarly in each case.

If you are wondering at this point whether we are really proposing that knowledge attribution may function in essentially the same way in nonhuman primates as it does in human infants and adults, let us be clear. We are. In fact, the commentary by **Moss** suggests that it might even extend to philosophers. Moss argues that the history of philosophy suggests that explicit theories of knowledge preceded those of belief in the Presocratics. As she argues, this suggests that explicitly theorizing about knowledge may be easier for creatures like us than explicit theorizing about belief. It may then be no coincidence that this empirical fact aligns with the other ones we reviewed in our target article and may provide yet another indicator that knowledge is more basic than belief for creatures like us – even those of us who are philosophers.

R4. Knowing what you don't know

A third objection that was touched on by a number of the commentaries was that theory of mind is for predicting and explaining behavior (see, e.g., **Binmore**, **Dudley, Gordon & Kovács**). This

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

perspective makes sense if one is committed to belief being the most basic theory of mind representation. But, if we are right that knowledge is more basic than belief, then the trouble faced by this approach is that the more basic form of theory of mind seems oddly ill-designed for action prediction and explanation, in particular (see **Bazhydai & Harris** for a similar line of reasoning). One notable feature of knowledge representation is that it seems to require more than justified true belief. But of course, justified true belief should be more than sufficient if your goal is just to predict someone's actions. Even unjustified false beliefs will do. A second notable feature of knowledge representation is that it allows you to represent others as knowing more than you yourself know. Others know all sorts of things you don't. But just knowing *that* others know more than you doesn't do you much good if your primary goal is predicting what they are going to do or explaining why they did what they did. So it's odd that our theory of mind capacity would have these particular features if it primarily evolved for the purpose of predicting and explaining behavior.

In contrast, if knowledge attributions involve representing others as understanding the actual world, then the ability to represent others as knowing more than you isn't particularly puzzling. In fact, it's precisely what you'd expect. When you represent someone as knowing more than you, you represent them as knowing something about the actual world you do not. You probably do not know how to play the zither, but you do think that there are, in fact, ways of playing the zither. And if you didn't think there was a fact of the matter, you couldn't represent someone as knowing that fact. For example, those of us who don't think each person's soul weighs a certain amount can't represent others as knowing the amount each soul weighs. Further, in cases where you yourself don't exactly know something, but you have a pretty good idea about it, you have a correspondingly good idea of what it is that the other person knows. And when you yourself have a great idea about the relevant part of the world, you'll have a correspondingly great idea about the content of someone else's mind. If you know why you randomly assign participants to conditions in a controlled experiment, and you represent someone else as knowing why too, then you'll have a great idea of exactly what it is they know. Not only do you know the precise content of their mental states, but you'll be able to predict what they'll do, and explain why they did what they did.

So it's not that we don't think knowledge representations can be used for prediction and explanation or that these representations don't reflect the content of others thoughts, it's just that the traditional view gets it backward. Knowledge concerns the actual world and which parts of it others understand. In some cases, others' understanding of the actual world will align with yours, and in those cases, you will know the content of others' thoughts, and be able to predict and explain their behavior. But there are also cases in which others' know more about the actual world than you do. Our point is that your own representation of the actual world plays much the same role both when you attribute knowledge to another of some fact you *do* know and when you attribute knowledge of facts you do *not* know. In both cases, you are representing another as understanding some part of the actual world (the way *that* part of the world actually is). What is changing is simply your own understanding of that part of the world (see **Durdevic & Krupenye** for related discussion).

This proposal for how to understand others as knowing more than you (egocentric ignorance) differs in important ways from the suggestions raised in many of the commentaries. For comparison, consider the proposal by **Tomasello** that the basic form of knowledge involves only knowledge-by-acquaintance. On this view, nonhuman primates only represent others as having been acquainted (or not) with physical objects in the world. Following **Kampis & Csibra**, suppose that the mechanism here works by simply tagging which physical objects someone is acquainted with. As Kampis & Csibra point out, such a mechanism does not seem to allow for representations of egocentric ignorance. To make this concrete, consider the success apes have in representing conspecifics as knowing whether there is a piece of fruit in a given box even when they themselves do not know (e.g., Kaminski, Call, & Tomasello, 2008). In such cases, subjects don't actually represent there being a piece of fruit in the box, and thus it's hard to see how they could tag that object as having been acquainted with the relevant conspecific. What this example illustrates is the difficulty in accounting for egocentric ignorance faced by views that reduce knowledge representations to simple representations like acquaintance, tagging, visual perspective, or perceptual access. Even more perplexing is how this kind of knowledge representation could be extended to understanding others as knowing how to crack open a nut (as nonhuman primates do, Rapaport & Brown, 2008) or knowing how to play the zither (as we humans do). What others know in such cases are not objects that can be tagged and clearly cannot be reduced to some particular visual perspective.

At the same time, our proposal for how to understand what happens when you represent others as knowing *less* than you (altercentric ignorance) also differs from those discussed in the commentaries. For example, **Deschrijver** suggests that altercentric ignorance may amount to simply attributing no representation whatsoever to another agent – much like the representation you attributed to the Prince of Liechtenstein before reading this sentence. But, just as it is possible to represent someone as sharing your knowledge of some particular part of the world (e.g., knowing a person) without representing them as sharing all your knowledge (e.g., knowing all the people you know of), it is possible to represent someone as *not* sharing your knowledge of a particular part of the world, without representing them as not sharing any of your knowledge. That is, the basic capacity for knowledge attribution allows for representations of knowledge and ignorance about specific parts of the world (this point provides a helpful contrast with the suggestion from **Gordon** that we may simply attribute all of our knowledge to others by default). In fact, much of the evidence we reviewed demonstrates precisely this kind of specificity in attributions of knowledge and ignorance. Consider simple studies in which nonhuman primates will selectively steal the piece of food that a dominant competitor does not know about (Hare, Call, Agnetta, & Tomasello, 2000). Success on these tasks requires that chimpanzees selectively represent the dominant competitor as ignorant of the existence of one piece of food while knowledgeable about the other. If they simply attributed no representation whatsoever to the other chimpanzee, they should be equally likely to take either piece of food (because the other chimpanzee would be equally unaware of both). When one represents others as ignorant, there must be specific parts of the world you do not represent them as knowing.

Importantly, the kind of ignorance we have been discussing does not involve representing someone else as being *aware* of their own ignorance (this would require a separate capacity involving meta-representation, see **Durdevic & Krupenye**). The difference is that when you represent another agent as being selectively ignorant about some part of the actual world, the

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

predictions you'll make concern only the other parts of the actual world they do know about (you'd predict that they'd be upset about you eating the food they do know about, but not the food that they don't know about). But, if you are able to represent other agents as having some awareness of their own ignorance (knowing *that* they don't know), then the predictions you'll make may also concern the ignorance itself, and what other agent's might do to alter their ignorance. As emphasized by **Royka & Jara-Ettinger**, one possibility is that such a meta-representation requires a mind that can employ some kind of symbolic negation operator, allowing you to represent the agent as knowing that they *do not* know. Future work may want to explore this possibility.

We have been arguing for an understanding of knowledge ascription that is both rich and flexible in some ways and notably limited in others. However, both the richness and the limits arise from a single unassuming commitment: When one represents others as knowing or not knowing something, one represents them as knowing or not knowing something about the actual world.

R5. Give learning a try

This way of understanding knowledge fits seamlessly with our claim that knowledge is for learning. When you represent your friend as knowing how to ride a bike, even though you don't know how to, you take them to understand something about the actual world: a way in which bikes can be ridden. You are not particularly interested in their ideas about how to ride a bike, independent of whether they actually work; what you take them to know and what you want them to teach you is how to actually ride a bike.

A number of commentaries pushed back on this basic idea, arguing that the capacity for belief representation is a better candidate for underwriting learning from others, especially given the success of cultural evolution in humans in particular (**Dudley, Gordon & Kovács**; **Richardson**; **Salazar**; **Sobel**).

One form of this objection was succinctly put by **Richardson**, who argues that knowledge cannot be both what humans share with nonhuman primates and what explains humans' unique capacity for cultural accumulation. Stated this way, we couldn't agree more. While we do think that the capacity for knowledge attribution is likely shared with nonhuman primates, we agree that knowledge is not what explains humans' unique capacity for cultural learning. Rather, we suspect that humans' unique cultural accumulation of knowledge is instead explained by our unique *representational capacities* – perhaps the capacity for representing abstract propositions, encoding information linguistically, and so on. We believe these sorts of capacities, not the capacity for knowledge attribution, are what differentiates humans from other species. But of course, none of this means that knowledge attribution doesn't play a central role in the process of accumulating cultural knowledge. If knowledge attribution works the way we've been arguing, then changes in domain-general representational capacities will result in changes in what we can and do represent others as knowing, which in turn will change what we can learn from them. And so, while nonhuman primates may accumulate knowledge of foraging techniques (e.g., Musgrave et al., 2020), human infants may accumulate knowledge of the names of novel objects (**Bazhydai & Harris**), and human adults may accumulate knowledge of math, all while using the same basic capacity for representing others as knowing something about the actual world.

While this response may help to address the differences in the content of cultural learning in nonhuman primates, there are clearly differences not only in content but also in frequency and tendency. We suspect that our proposal has little to contribute in explaining these differences. There are myriad ways in which humans are both more social and more successful in communicating than our primate relatives (see Henrich, 2015 for a discussion).

A second form of the objection that knowledge is for learning, raised by **Sobel** among others, is that the processes for selectively determining who to learn from may be better accounted for by a form of belief representation. In a helpful response, the commentary by **Bazhydai & Harris** provides a beautiful accounting of the empirical evidence that knowledge rather than belief representations support selective learning in infants. The body of work they discuss demonstrates that infants selectively learn from others who are knowledgeable and selectively pass on information to those who are ignorant, all while not yet demonstrating any real capacity for the kind of meta-representation required by belief. This literature similarly helps to address the point raised by both **Handley-Miner & Young** and **Kampis & Csibra** that if knowledge representations are going to be useful for social learning, they need to be accompanied by mechanisms for determining who actually knows what you want to know. The literature on trust in testimony provides remarkably thorough evidence for how these mechanisms may function, and we hope that this literature will become increasingly integrated into theory of mind research (see **Bazhydai & Harris**, **Salazar**, and Harris, Koenig, Corriveau, & Jaswal, 2018). On a related note, one would expect that what we choose to teach others is also guided by knowledge, and here again, there is a growing body of evidence that knowledge plays a key role in guiding the information we provide to others (Turri, 2016b).

Finally, it may be worth being explicit that none of this means that belief attribution cannot also support social learning. However, if we are correct about the essential difference between knowledge and belief, then the cases in which belief attribution plays an essential role will be ones in which what you need to learn is something specifically about how others think, not how the world actually is. When the emperor wears no clothes, successfully predicting and coordinating with others certainly will require belief-based social learning. Still, we suspect such cases make up the periphery rather than the core of learning from others, especially in the course of primate evolution.

R6. What to do with belief?

Throughout, we have been arguing for a central way of understanding the differences between knowledge and belief attribution. An important separate question, which was raised in a number of commentaries, instead asks how these two forms of attribution may be related to one another (**Bender & Gatewood, Brakel, Durdevic & Krupenye, Kano & Call, Nagel, Ninan**).

As pointed out in the commentary by **Nagel**, the account of knowledge attribution we've given can occur entirely independently from belief attribution. Thus, our view differs in an important way from standard philosophical views of knowledge, according to which knowledge entails belief. It is important here to keep in mind the difference between the philosophical claim about the concept of knowledge and the psychological

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

claim we have made (see the commentary by **Gerken** for a similar concern). We've argued that the representation of knowledge does not entail the representation of belief. This point is directly supported by the empirical evidence. For example, we've argued that monkeys can represent knowledge but not belief. And if that's right, it clearly can't be the case that their representing knowledge entails their representing belief. This same point is also supported by the growing experimental philosophy evidence for cases in which people will attribute knowledge but not belief (Myers-Schulz & Schwitzgebel, 2013; Yuan & Kim, forthcoming). Such cases can naturally be described as ones in which we take the agent to have access to the relevant part of the world even though this access doesn't exhibit the normal impact on the agent's thoughts or behavior (see the commentary by **Brakel** for related ideas).

One way of thinking about this independence between knowledge and belief aligns with the proposal from **Kano & Call**, according to which the capacities for knowledge and belief attribution are entirely separate. Kano & Call agree that human infants seem to have an ability to attribute knowledge but not belief. However, they are moved by the studies providing evidence for false-belief representation in apes (Kano, Krupenye, Hirata, Tomonaga, & Call, 2019; Krupenye, Kano, Hirata, Call, & Tomasello, 2016; see also the commentary by **Durdevic & Krupenye**). Thus, as they argue, given that the two capacities do not consistently appear together, perhaps they should simply be understood as arising from separate systems. While we are less convinced that the existing research provides sufficient evidence for a capacity for belief representation in nonhuman primates (and monkeys in particular, see sect. 4.1), we can set this question aside for now. Note that if Kano and Call turn out to be correct that some nonhuman primates have a capacity for belief representation, there would remain a remarkably consistent pattern across species: One never finds a capacity to represent beliefs in the absence of a capacity to represent knowledge.

This consistent pattern suggests an alternative way of understanding the relation between knowledge and belief that aligns instead with the proposal from **Ninan**. As argued by Ninan, the capacity for belief representation may depend on a prior ability to represent knowledge. Following an idea from Williamson (2002), Ninan suggests that instances of representing others as believing something may essentially be instances of representing someone as acting *as if they knew something*. If this is right, then belief attribution (even in cases where the belief is true), would require a form of counterfactual conditional reasoning. In other words, it would require representing a merely possible way the actual world could have been, and then taking the agent to be related to that world in much the same way we take others to be related to the actual world when they know things about it. Three features make Ninan's proposal intriguing. The first is that it could explain the general pattern whereby belief attribution appears later in development than knowledge attribution. The second is that it fits well with the empirical correspondence one finds in human development between counterfactual conditional reasoning and belief attribution (Riggs & Peterson, 2000). And the third is that it provides one way of understanding why there are cases in which knowledge does not entail belief, since knowing something does not entail acting as if one knew that thing (Myers-Schulz & Schwitzgebel, 2013; Radford, 1966).

While we are not yet sure whether belief representation should be understood as depending in some way on knowledge representation, this is clearly an important area for future research.

R7. One thousand flowers

There remains a great deal we do not know about the basic theory of mind capacity we have been concerned with. At least partially, this is because knowledge attribution has received comparatively less attention than belief attribution in the history of theory of mind research. So, while we agree with **Dudley & Kovács**, **Kano & Call**, and **Richardson** that our paper should probably not incite a wholesale abandonment of the study of belief attribution, we want to emphasize the range of commentaries that pointed to important new questions and future directions for the study of knowledge. We hope these questions spur a new era in theory of mind research.

R7.1. Catching up

In the past 40-plus years – starting from the proposal of the false-belief task in the commentaries to Premack & Woodruff's (1978) article in this journal – we have learned a great deal about belief representation. We have largely reached a consensus on the neural substrates involved in representing false beliefs (e.g., Saxe and Kanwisher, 2003). We have developed elegant ways of computationally modeling the process of belief attribution and update (e.g., Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017; Jara-Ettinger, Schulz, & Tenenbaum, 2020). We have an increasingly good idea of when the capacity for belief attribution arose over the course of evolution (e.g., Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011). And, we have thoroughly studied the extent to which humans automatically represent others' false beliefs (e.g., Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Kovács, Téglás, & Endress, 2010; Phillips et al., 2015). Yet, as pointed out in many of the commentaries, we have corresponding gaps about each of these when it comes to knowledge.

R7.1.1. The neuroscience of knowledge attribution

The commentaries by **Bricker** and **Gordon** call for the emergence of the neuroscientific study of knowledge attribution. Bricker's EEG study (Bricker, 2020) is a helpful early step in this direction. He finds that belief representation demands more neural resources than knowledge representation as indicated by differences in P3b amplitude. A clear implication of this finding is that knowledge representation – even propositional knowledge representation – does not depend on belief representation, since representing the agent's knowledge requires less processing than representing the agent's beliefs. Still, many open questions remain. Because theory of mind networks have quite literally been defined by false beliefs (i.e., a false belief vs. false photograph contrast, Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011), we don't yet have much of an understanding of the neural mechanisms involved in knowledge representation. Thus, an important and completely open question is whether knowledge representation recruits the same theory of mind network as belief representation or relies on a distinct set of neural substrates.

R7.1.2. The computation of knowledge attribution

The commentaries by **Asaba et al.**, **Durdevic & Krupenye**, and **Royka & Jara-Ettinger** emphasize the importance of understanding the computational processes that underwrite knowledge attribution. The existing research on computational theory of mind

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

relies on inferences over belief states, whether through Bayesian inference (Baker et al., 2017), inverse reinforcement learning (Jara-Ettinger et al., 2020), or another mechanism (Koster-Hale & Saxe, 2013). What the current proposal suggests is that there may be simpler forms of theory of mind computation that do not require representing and reasoning over the potentially huge number of wrong beliefs an agent may have. Moreover, if the current proposal is correct, then the computations that underwrite knowledge attribution may instead directly recruit one's own understanding of the world, which would serve to drastically reduce the space of possible knowledge states necessary to reason over. We hope future work takes up this challenge.

R7.1.3. The evolution of knowledge attribution

While we've argued that knowledge arose before belief, this does not settle the question of *when* the capacity for knowledge representation actually evolved. The commentary by **Veit** suggests this capacity arose during the Cambrian explosion. Maybe so, but either way, this is an empirical and testable claim we hope is taken up in future work by studying knowledge representations in species less related to us than nonhuman primates, such as corvids, canines, or even octopuses.

R7.1.4. The automaticity of knowledge attribution

The commentary by **Surtees & Todd** points out that a great deal remains to be done in studying implicit, spontaneous, or automatic knowledge representations. As we've argued, and was expanded on by Surtees & Todd, most of the current evidence is based on visual perspective taking tasks, which at best can only provide suggestive evidence for the automatic calculation of genuine knowledge representations. (The evidence is less equivocal about belief representations, which clearly do not seem to be automatic.) While there have been a few studies that have looked at abstract knowledge rather than visual perspective taking (e.g., Dungan & Saxe, 2012), the question of whether we automatically calculate what others know, and what the limits of these calculations are, remain important questions for future work.

R7.2. Looking forward

In addition to commentaries proposing that we need to understand knowledge in the same ways we've come to understand belief, other commentaries emphasized that there are aspects of knowledge that merit studying on their own grounds.

In this vein, the commentary by **Gerken** points toward the importance of studying the biases and limits of knowledge representation. As Gerken argues, some interesting features of knowledge representations may provide further clues to how this capacity functions. We agree that studying the signature limits of knowledge ascription is an important and productive avenue for future work. We suspect that this approach will also help uncover ways in which knowledge and true belief attribution come apart, the importance of which was emphasized by **Lassiter** and **Durdevic & Krupenye**.

Similarly, the commentary by **Machery, Barrett, & Stich (Machery et al.)** argues for the importance of studying cross-cultural and cross-linguistic variation in knowledge ascription (also emphasized by **Bender & Gatewood**). While it would be surprising if there was genuinely no cross-cultural or cross-linguistic variation in knowledge ascription, the extant evidence indicates that many of the notable features of knowledge attribution exhibit remarkable cross-cultural stability. For example, a well-known developmental finding is that there is remarkable stability in the order in which children pass a battery of theory of mind tasks (Wellman & Liu, 2004), and variations across languages and cultures are relatively minor (e.g., Shahaeian, Peterson, Slaughter, & Wellman, 2011). Moreover, Machery et al. (and their colleagues) have found robust evidence that knowledge is denied across cultures in Gettier cases (Machery et al., 2017) and that knowledge ascriptions are equally insensitive to stakes across cultures (Rose et al., 2019). And there is even new evidence for cross-cultural stability in the tendency to attribute knowledge in cases where belief is denied (Yuan & Kim, forthcoming). In their commentary, Machery et al. hint at some preliminary evidence that there may be cases in which this last feature of knowledge is not exhibited. If those results hold up, it would certainly be interesting and important. Still, we do not think that such a finding by itself would be problematic for our general proposal. If the capacity for knowledge representation is indeed basic in the way we've argued, one should expect a lot of generality across languages and cultures, but probably not strict universality (see Strickland, 2017). More importantly though, the only way to know whether this generality claim holds up is to do the difficult and important cross-cultural work being done by Machery et al. Thus, we echo their call to continue investigating cross-cultural and cross-linguistic variation in knowledge attribution.

A final group of commentaries emphasized the importance of better understanding knowledge representations in our social lives, especially in cases in which we interact with and learn from others.

The commentary by **Bazhydai & Harris** calls for studying the relationship between knowledge representation and active solicitation of teaching from and to others. As they emphasize, an important but as of yet unanswered question is whether young children exhibit higher rates of soliciting information from others when they represent them as knowing something rather than (merely) truly believing it. In a similar vein, the commentary by **Erdemli, Audrin, & Sander** suggests that the process of learning from others may partially be driven by "social epistemic emotions" and "affective social learning." We hope that researchers working on active solicitation begin to research these important questions.

Relatedly, **Asaba et al.** and **Handley-Miner & Young** emphasize the importance of studying knowledge representation in cases of real-world complexity, where people may only have partial knowledge and you may even be uncertain about who has knowledge or how much knowledge they have. Following much of the empirical research, we have emphasized cases where knowledge is relatively clear-cut. However, many real-world cases involve precisely the kind of uncertainty Handley-Miner and Young point out. The literature on trust in testimony provides a rich resource to draw on (see Harris et al. 2018 for a recent review), but better understanding knowledge attribution in the face of such uncertainty clearly remains an important avenue for future work.

Asaba et al., Schlicht et al., and **Woo, Tan, & Hamlin (Woo et al.)** all raise important questions concerning theory of mind about others' goals or preferences. An ability to represent others' goals and preferences, much like the ability to represent knowledge, appears early in development and before belief representation (see the commentaries by Schlicht et al. and Woo et al.). Notice that when you represent others as having goals or preferences, these seem to involve the actual world. Others may have a goal of getting to a particular part of the actual world (say, the top of a hill), or a preference for eating some part of the world (say, cookies). An intriguing possibility then is that this form of theory of mind, much like knowledge, essentially involves one's

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press

own understanding of the actual world. And if this is correct, then we would *not* expect an early facility in attributing goals or desires, when the object of those goals or desires is precluded by the actual world (e.g., wanting to eat a cookie now that was already eaten yesterday). We hope future work investigates this possibility.

And with that, let us turn to a new chapter in theory of mind research.

Notes

1. We'd like to note that some of the authors (WB and JT) have recently challenged views on which the factivity of knowledge requires that one can only know things that are strictly speaking or precisely true (Buckwalter & Turri, 2020a, 2020b).

References

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1–10.

Bricker, A. M. (2020). The neural and cognitive mechanisms of knowledge attribution: An EEG study. *Cognition*, *203*, 104412.

Buckwalter, W., & Turri, J. (2020a). Knowledge and truth: A skeptical challenge. *Pacific Philosophical Quarterly*, *101*(1), 93–101.

Buckwalter, W., & Turri, J. (2020b). Knowledge, adequacy, and approximate truth. *Consciousness and Cognition*, *83*, 102950.

Colaço, D., Buckwalter, W., Stich, S., & Machery, E. (2014). Epistemic intuitions in fake-barn thought experiments. *Episteme; Rivista Critica di Storia Delle Scienze Mediche e Biologiche*, *11*(2), 199–212.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *Neuroimage*, *55*(2), 705–712.

Dungan, J., & Saxe, R. (2012). Matched false-belief performance during verbal and nonverbal interference. *Cognitive Science*, *36*, 1148–1156. doi: 10.1111/j.1551-6709.2012.01248.x.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, *79*(5), 836–848.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785.

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273.

Henrich, J. (2015). *The secret of our success*. Princeton University Press.

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*, 72–80.

Horschler, D. J., Santos, L. R., & MacLean, E. L. (2021). How do non-human primates represent others' awareness of where objects are hidden? *Cognition*, *212*, 104658.

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naive utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences*, *116*, 20904–20909. doi: 201910095. https://doi.org/10.1073/pnas.1910095116.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114, https://doi.org/10.1126/science.aaf8110.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*(4), 412.

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., … Hashimoto, T. (2017). Gettier across cultures 1. *Noûs*, *51*(3), 645–664.

Marticorena, D. C., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science*, *14*(6), 1406–1416.

Musgrave, S., Lonsdorf, E., Morgan, D., Prestipino, M., Bernstein-Kurtycz, L., Mundry, R., & Sanz, C. (2020). Teaching varies with task complexity in wild chimpanzees. *Proceedings of the National Academy of Sciences*, *117*(2), 969–976.

Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing that P without believing that P. *Noûs*, *47*(2), 371–384.

Phillips, J., Knobe, J., Strickland, B., Armary, P., & Cushman, F. (2018). Evidence for evaluations of knowledge prior to belief. In Proceedings of the *Cognitive Science Society*.

Phillips, J., Ong, D. C., Surtees, A. D., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A second look at automatic theory of mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*, *26*(9), 1353–1367.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Radford, C. (1966). Knowledge: By examples. *Analysis*, *27*(1), 1–11.

Rapaport, L. G., & Brown, G. R. (2008). Social influences on foraging behavior in young nonhuman primates: Learning what, where, and how to eat. *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(4), 189–201.

Riggs, K. J., & Peterson, D. M. (2000). Counterfactual thinking in pre-school children: Mental state and causal inferences. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind*, (pp. 87–99). Psychology Press/Taylor & Francis (UK).

Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., … Zhu, J. (2019). Nothing at stake in knowledge. *Noûs*, *53*(1), 224–247.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *Neuroimage*, *19*(4), 1835–1842.

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*(5), 1239.

Strickland, B. (2017). Language reflects "core" cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive Science*, *41*(1), 70–101.

Turri, J. (2016a). A new paradigm for epistemology: From reliabilism to abilism. *Ergo*, *3*(8), 189–231.

Turri, J. (2016b). *Knowledge and the norm of assertion: An essay in philosophical science*. Open Book Publishers.

Turri, J. (2018). Primate social cognition and the core human knowledge concept. In M. Mizumoto, S. Stich, & E. McCready (Eds.), *Epistemology for the rest of the world: Linguistic and cultural diversity and epistemology*, (pp. 279–290). Oxford University Press.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.

Williamson, T. (2002). *Knowledge and its limits*. Oxford University Press on Demand.

Yuan, Y., & Kim, M. (forthcoming). Cross-Cultural universality of knowledge attributions. *Review of Philosophy and Psychology*.

https://doi.org/10.1017/S0140525X20000618 Published online by Cambridge University Press