

## Behavioral and Brain Sciences (forthcoming)

---

**This Target Article has been accepted for publication and has not yet been copyedited and proofread. The article may be cited using its doi (About doi), but it must be made clear that it is not the final version.**

Long abstract: 250 words  
Main text: 12641 words  
References: 5243 words  
Total: 18427

### Knowledge before Belief

Jonathan Phillips (corresponding author)  
Program in Cognitive Science, Dartmouth College  
Email: [jonathan.s.phillips@dartmouth.edu](mailto:jonathan.s.phillips@dartmouth.edu)  
Homepage: <https://www.dartmouth.edu/~phillab/phillips.html>

Wesley Buckwalter  
Department of Philosophy; School of Social Sciences, University of Manchester  
Email: [wesleybuckwalter@gmail.com](mailto:wesleybuckwalter@gmail.com)  
Homepage: <https://wesleybuckwalter.org/>

Fiery Cushman  
Department of Psychology, Harvard University  
Email: [cushman@fas.harvard.edu](mailto:cushman@fas.harvard.edu)  
Homepage: <http://cushmanlab.fas.harvard.edu/>

Ori Friedman  
Department of Psychology, University of Waterloo  
Email: [friedman@uwaterloo.ca](mailto:friedman@uwaterloo.ca)  
Homepage: <https://sites.google.com/view/uwaterloocclab>

Alia Martin  
School of Psychology, Victoria University of Wellington  
Email: [alia.martin@vuw.ac.nz](mailto:alia.martin@vuw.ac.nz)  
Homepage: <https://vuwbabylab.com/>

John Turri  
Philosophy Department and Cognitive Science Program, University of Waterloo  
Email: [john.turri@gmail.com](mailto:john.turri@gmail.com)  
Homepage: <https://john.turri.org/>

Laurie Santos  
Department of Psychology, Yale University  
Email: [laurie.santos@yale.edu](mailto:laurie.santos@yale.edu)  
Homepage: <https://caplab.yale.edu/>

Joshua Knobe

Program in Cognitive Science; Department of Philosophy, Yale University

Email: [joshua.knobe@yale.edu](mailto:joshua.knobe@yale.edu)

Homepage: <https://campuspress.yale.edu/joshuaknobe/>

Short abstract:

Research on the capacity to understand others' minds has tended to focus on *belief*, which is thought to be among the most central and basic theory-of-mind representations. Representations of *knowledge*, by contrast, have received comparatively little attention and are often thought to rely on prior belief representations. We review evidence from a wide range of fields employing diverse methodological approaches including comparative psychology, developmental psychology, social psychology, and experimental philosophy. By every measure, the evidence indicates that representations of knowledge are more basic than representations of beliefs. We argue that this calls for a shift in focus in theory-of-mind research.

Long 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 doesn't 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 non-human 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 exhibit 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.

Keywords:

belief, factivity, false belief, knowledge, knowledge-first, theory of mind

## 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.<sup>1</sup>

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. Clearly, 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 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 work 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 non-human primates attribute knowledge but not belief (§4.1), that the capacity to attribute knowledge arises earlier in human development than the capacity to attribute belief does (§4.2), that implicit knowledge attributions are made more automatically than belief attributions (§4.3), that the capacity to represent knowledge may remain intact in patient populations, even

---

<sup>1</sup> 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.

when belief representation is disrupted (§4.4), and that explicit knowledge attributions do not depend on belief attributions (§5.1-2) and are made more quickly than belief attributions (§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:

### (i) Knowledge is factive.

You can believe whatever you like, but you can only know things that are true. And when you represent others' knowledge, you can only represent them as knowing things that you take to be true. If you don't think it's true that the moon landing was faked, you can't rightly say, “Bob knows the moon landing was faked”. You'd 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 (Kiparksy & Kiparksy, 1970; Williamson, 2000).

### (ii) 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 kind 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; Macherly, et al., 2017; Starmans & Friedman, 2012).

### (iii) Others can know things you don't.

While you can't represent others as knowing things that are false, you can represent others as knowing things you don't know. You can say, for example, "Suzy knows where to buy an Italian newspaper" even if you yourself don't 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 yourself do ("egocentric ignorance") (Nagel, 2017; Phillips & Norby, 2019).

(iv) 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, e.g., 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, while representations of seeing or hearing, unlike knowledge, are modality-specific. Thus, 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'll 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 work 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 other (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 gives 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, et al., 2017; Goldman, 2006; Gopnik & Wellman, 1992; Gordon, 1986; Leslie, et al., 2004). Then, to determine whether or not John *knows* that the keys are in his pocket, we'd 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 for 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, et al., 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.

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 don't think that it's true that John's keys are in his pocket, then you certainly can't represent him as knowing that his keys are in his pocket. Moreover, even if you do think it's 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 can't represent him as knowing that his keys are in his pocket.

By 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. Thus, 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, i.e., belief that has various further properties (being true, justified, etc.). 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, et al., 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 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 non-linguistic 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 a 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, ASD, etc.). 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 at work, 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 which require the kinds of complex computational abilities that only adult humans possess. But which parts of this numerical understanding are basic and which 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 et al. 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, et al., 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, et al., 2003). Finally, researchers have identified special populations in which participants’ understanding of approximate numbers are preserved despite radical differences in experience (e.g., the Mundurucu, an indigenous group of people who live in the Amazon River basin, Dehaene, et al., 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 work 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 & Belief in Nonhuman Primates

The first empirical tool that we marshal is comparative work examining what non-human 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 work 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 non-human 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, et al., 2017; Kano, et al., 2019; Krupenye, et al., 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, et al., 2017; Kano, et al., 2019; Krupenye, et al., 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, et al., 2018; Kulke, et al., 2018; Schuwerk, et al., 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, et al., 2008; Krachun, et al., 2009; O'Connell & Dunbar, 2003). In one study for example, Kaminski and colleagues (2008) explored whether chimpanzees could understand that a competitor had a false belief about the location of a hidden food item. Kaminski and colleagues used a design in which subject chimpanzees competed with a conspecific for access to contested foods (for the first of such tasks, see Hare, et al., 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 the 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 (Brauer, et al., 2007; Hare et al. 2000; Hare, et al., 2001; Hare, et al., 2006; Kaminski et al., 2008; Karg, et al., 2015; Krachun et al., 2009; Melis, et al., 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 non-human 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 understand other individuals' beliefs, even when tested on tasks that human infants have passed (Marticorena, et al., 2011; Martin & Santos, 2016). Work 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 move 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 into 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 with 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 can't see the food (Flombaum & Santos, 2005) or who can't 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 watch an object go into one of two locations, they look longer when that person reaches to 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 non-human primates actually represent *knowledge*?

An essential further question is whether the research on non-human 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 non-

human 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 non-human primates can represent egocentric ignorance; that is, they can represent someone else as knowing something they themselves 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 themselves 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<sup>2</sup>. Critically, in a minimally different false-belief condition where the containers switched positions while 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 (Kaminski et al., 2008; Horschler, et al., 2019). Specifically, these studies included conditions where food was placed in one of two opaque containers in clear sight of both the experimenter and the non-human 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, in these conditions, non-human 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 (Kaminski et al., 2008, Horschler, et al., 2019).

Finally, there is evidence that the knowledge representations found in non-human 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 work also suggests 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 work to date suggests that the capacity to think about others' knowledge predates an ability to represent others' beliefs.

## 4.2 Knowledge and Belief in Human Development

---

<sup>2</sup> 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.

### 4.2.1. Knowledge and belief in infancy

Just as studies of non-human primates can provide evidence about which cognitive capacities are evolutionarily more foundational, so too can studies of pre-verbal 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 for 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* by 15 months of age, or earlier (Buttelmann, et al., 2009; Kovács et al., 2010; Onishi & Baillargeon, 2005; Surian, et al., 2007; Träuble, et al., 2010). At the same time, there have been a number of compelling plausible proposals for how key behavioral patterns can be explained without a genuine ability for belief representation (Priewasser, et al., 2018; Butterfill & Apperly, 2013; Burge, 2018; Heyes, 2014a). 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, et al., 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 6 months of age. First, 6-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, 6-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. But 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 hasn't seen both options. In this way, infants recognize that an agent's knowledge of her surroundings affects that 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). Thus, six-month-olds seem to recognize some of the conditions under which an individual will become knowledgeable about information (Vouloumanos, et al., 2014). These two examples and many others (e.g., Hamlin, et al., 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 (though see Kovács et al., 2010).

An important aspect of these studies (as well as the non-human primate research) is that they required researchers to rely on tasks with solely *non-linguistic responses*. To complement this work, we next turn to considering 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, et al., 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, et al., 1989; Shatz, et al., 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, et al., 1989). Moreover, children make systematic errors when using non-factive 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, et al., 2012; Sowalsky, et al., 2009, though see Dudley, et al., 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, e.g., the belief is false. That is, when children are asked whether an agent “thinks” something, their pattern 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, et al., 2017; Harris, et al., 2017; Harris, et al., 2017).

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 this, 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 while children are aged 3 (e.g., Pillow, 1989; Pratt & Bryant, 1990; Woolley & Wellman, 1993), successful attribution of false belief typically occurs only when they are 4 or older (Grosse Wiesmann, et al., 2017; see Wellman, et al., 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, et al., 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, et al., 2005), and children from non-Western cultures (Shahaeian, et al., 2011; Shahaeian, et al., 2014; Wellman, et al., 2006).

Considering young children's capacity for making explicit, verbal judgments of knowledge and belief, one sees a familiar pattern emerge. Much like in nonverbal tasks, young children succeed on 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 themselves don't? Behne and colleagues (2012) had an experimenter hide an object in one of 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 and colleagues (2014) provided evidence that 12-month-old infants' own pointing is used to query others who they take to be more knowledgeable than they themselves 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 & 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 work provides evidence that infants have an early-emerging capacity to represent others as knowing something that they themselves 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, et al., 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, et al., 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 yourself 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, et al., 2010; Fabricius & Imbens-Bailey, 2000; Oktay-Gür & Rakoczy, 2017; Perner, et al., 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, et al., 2012; Moll, et al., 2014). Additionally, infants attribute knowledge based on non-perceptual inferences that the agent should make. For example, Träuble and colleagues (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, et al., 1992; Papafragou, et al., 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.

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 don’t

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 3 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 are not automatic. Representations of 3 work differently. The processes involved in representations of 3 operate *automatically*: you couldn't *not* realize that there are 3 dots, even if you wanted to (Dehaene & Cohen, 1994; Kaufman, et al., 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, while at the same time, systematically varying the information that was presented to another (irrelevant) agent in the experiment (Kovács et al., 2010; Samson, et al., 2010; Surtees, et al., 2016; Low & Watts, 2013, Apperly & Butterfill, 2009). Researchers could then ask whether participants were automatically (i.e., mandatorily) calculating the mental states of the irrelevant agent by asking whether participants' own 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, et al., 2018; El Kaddouri, et al., 2019; van der Wel, et al., 2014), and considerable evidence that strongly suggests that belief representation is *not* automatic (Apperly, et al., 2006; Kulke, et al., 2019; Low & Edwards, 2018; Surtees, et al., 2012; Surtees, et al., 2016).

Complementary evidence comes from studies asking whether representations of others' beliefs are computed when attentional resources or executive function are taxed. This body of research provides clear evidence that representing what others believe requires deliberative attention and executive function (Apperly, et al., 2008; Apperly, et al., 2009; Dungan & Saxe, 2012; Qureshi, et al., 2010; Schneider, et al., 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 while 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 than representations of 3.

#### 4.3.2. Current evidence for automatic knowledge representation

By contrast, one finds intriguing evidence that human adults may automatically represent what others know. In one study, Samson and colleagues (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 own 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., 2016) 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 continuing debate about whether these findings reflect 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, et al., 2017; Santiesteban, et al., 2014), and others providing empirical evidence against the proposed confounds (Furlanetto, et al., 2016; Gardner, et al., 2018; Marshall, et al., 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 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 automatic theory of mind

The research reviewed in the previous sections on non-human 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 non-human primate relatives and early emerging in human development—also tend to operate automatically in humans. Thus, 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 for you to represent someone else as 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 work on automatic theory of mind, and thus point to exciting new avenues for future work as research on this topic presses forward.

#### 4.4 Evidence from patient populations

Another tool 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 studied in theory of mind research are people with Autism Spectrum Disorder (ASD). Research looking at theory of mind in people with ASD has found that they often have difficulties in correctly representing others' beliefs (Baron-Cohen, 1997; Baron-Cohen, et al., 1985; Frith, 2001; Moran, et al., 2011; Schneider, et al., 2013; Senju, et al., 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, et al., 1989; Pratt & Bryant, 1990). For example, in studies that tracked participants' eye-movements during true and false belief tasks, researchers found that while the eye movements of people with ASD differ from controls when the agent has a false belief, they actually don't 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, et al., 1997; Reed & Peterson, 1990; Tan & Harris, 1991). By 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 well than typically developing controls (Hamilton, et al., 2009; Leslie & Frith, 1988; Reed & Peterson, 1990; Yirmiya, et al., 1994).

Taken together, this research suggests that the capacity for representing others' beliefs is disrupted in patient populations, while 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 theory of mind. Primate work 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 automatic theory of mind in adult humans suggest that representations of knowledge may happen more automatically and effortlessly than representations of beliefs. And evidence from patient populations demonstrates that an ability to represent beliefs can be disrupted while 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 we've focused on 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 work has been conducted in the field of experimental philosophy. While we review the evidence in detail below, the lesson that comes from this work 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 (§5.1), that knowledge attributions sometimes occur when belief attributions do not (§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 (§5.3), and that knowledge attributions are better predictors of behavior than attributions of belief (§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 §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 work 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 for “think” is *more* frequent than the term for “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 vs. 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 doesn’t even believe it?

Surprisingly, however, results from experimental philosophy provide 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, et al., 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 suggests 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 for (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, et al., 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 relies primarily on which beliefs we attribute (e.g. Baker, et al., 2009; Rakoczy, 2009; Tomasello, et al., 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 work 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; Turri, 2015b; Turri, et al., 2017), these findings suggest a previously unrecognized role for knowledge attributions in predicting how people *will* behave.

## 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 work (§4.1), developmental work (§4.2), automatic processing in human adults (§4.3), and work with patient populations (§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 (§5). Critically, we've also illustrated that the theory of mind capacity revealed by these various methods exhibits a set of signature features that are specific to knowledge (§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 (§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. And 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. *Why* someone believes what they do seems entirely irrelevant for the purpose of 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 yourself do not know. It's easy to see why this isn't particularly useful for action prediction. Imagine a ball was placed in one of two boxes, but you don't 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 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 own 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 (Hintikka, 1975; Egré, 2008). 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've provided clear evidence that we can and do (§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've reviewed on non-human 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 non-human 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 & Brown, 2008) or how to solve novel problems (Call & Tomasello, 1995; Call, et al., 2005; 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, e.g., 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 (§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 non-human primates do (§4.2), then you have a capacity that is well-suited for helping you learn from others.

In a similar way, 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 and colleagues (2018), an impressive body of research has documented the capacity to understand others as sources of unknown 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, et al., 2018; Moses, et al., 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 8-months-old (Begus & Southgate, 2012; Brooker & Poulin-Dubois, 2013; Harris & Corriveau, 2014; Koenig & Harris, 2005; Poulin-Dubois & Brosseau-Liard, 2016; Tummeltshammer, et al., 2014; Zmyj, Buttelmann, et al., 2010). As reviewed above (§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, however, it seems obvious: when parents teach their children some fact 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've 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, et al., 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 (§4.2).

### 6.3. A Call to Arms

Since the 1970's, 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 of 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.

## **Acknowledgements**

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

## **Funding statement**

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).

**Conflicts of interest: none**

## References

- Apperly, I. A. (2010). *Mindreaders: The Cognitive Basis of "Theory of Mind."* Psychology Press, New York, NY. <http://doi.org/10.4324/9780203833926>
- 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., 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., 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, UK: 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*, 1–10. <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 and S. Lea (eds.) *The Descent of Mind: Psychological Perspectives on Hominid Evolution*. Oxford University Press. New York, NY. 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: 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: 359–75.
- Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, 1(4):557–60.
- 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*. Chicago: University of Chicago Press.
- 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 (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. (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.
- Carter, Gordon & Jarvis (eds.). (2017). *Knowledge-First: Approaches in Epistemology and Mind*. Oxford: 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>
- 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: Oxford University Press. doi:10.1093/0198238797.001.0001
- de Villiers, J. (1995). Questioning minds and answering machines. In D. MacLaughlin and S. McEwen (Eds.) *Proceedings of the Boston University Conference on Language Development*. 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*. Hove, U.K.: 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*. Somerville, MA: Cascadilla Press.
- 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):568–70
- Dennett, D. (1989). *The Intentional Stance*. Cambridge, MA: MIT 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. 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.
- Drayton, L., & Santos, L.R. (2018). What do monkeys know about others' knowledge?. *Cognition*, 170, 201-208.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: 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, Cham.
- Dungan, J. and 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
- Egre, P. (2008). Question-embedding and factivity. In F. Lihoreau (Ed.), *Grazer Philosophische Studien*, 77, 85–125.
- El Kaddouri, R., Bardi, L., De Breaecker, 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*, 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 and K. J. Riggs (Eds) *Children's Reasoning about The Mind* (pp. 267-280). Psychology Press, East Sussex, UK.
- 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). Lincoln: 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). Hillsdale, NJ: 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*. New York, NY: 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. (2017). 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., & 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.
- Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59, 771–785.
- Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*, 1(4): 576–77.
- 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.
- Harris, P.L. & Corriveau, K.H. (2014). Learning from testimony about religion and science. In *Trust and Skepticism: Children's Selective Learning from Testimony*, ed. E Robinson, S Einav, pp. 28–41. Psychology Press, Hove, UK
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, 69.
- Harris, P. L., Ronfard, S., & Bartz, D. (2017). 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. (2017). '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*. 1-25.
- 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 Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition) 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*, 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.
- 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.
- 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.
- 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, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy*, 1, 3-44.
- Kaufman, E.L., Lord, M.W., Reese, T.W., & Volkman, 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.
- Kiparsky, P. & Kiparsky, C. (1970). Fact. In M. Bierwisch and K. Heidolph (Eds.), *Progress in linguistics: A collection of papers* (pp. 143-173). Mouton & Co. N.V. Publishers, The Hague.
- 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 *The Routledge Handbook of Social Epistemology*, ed. M. Fricker, P. Graham, D. Henderson, N. Pederson, J. Wyatt. Routledge, New York, NY.
- 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: susceptibility 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.
- 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.
- 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., 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. In A. Chereches (Ed) *Proceedings of SALT*, 22, 247–267.
- Locke, J. (1689/1975). *An Essay Concerning Human Understanding*. P.H. Nidditch (ed.). Oxford: 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., Sirker, S., Usui, 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.
- Martcorena, 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.
- 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*. New York, NY: Oxford University Press.
- Nozick, R. (1981). *Philosophical Explanations*, Cambridge, MA: 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.
- 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.
- Oktaý-Gür, N., & Rakoczy, H. (2017). Children's difficulty with true belief tasks: Competence deficit or performance problem? *Cognition*, 166, 28-41.
- 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. New York, NY.
- 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., 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. & Norby, A. (2019). Factive theory of mind. *Mind & Language*, 1-24.  
<https://doi.org/10.1111/mila.12267>
- 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*. Indianapolis: 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.

- 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.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, 1(4), 515-526.
- 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*, 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*, 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: 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. D., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, 83(2), 452-460.
- Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, 150, 43-52.
- 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
- Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, 148, 97-105.
- 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*. Cambridge, MA: 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., and 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*, 1–22. 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: 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.