Cognition 270 (2026) 106400

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Exploring the evolutionary roots of theory of mind: Primate errors on false belief tasks reveal representational limits

Amanda Royka [*,1], Daniel J. Horschler [1], Walker Bargmann, Laurie R. Santos

Department of Psychology, Yale University, United States of America

ARTICLE INFO

Keywords:
Theory of mind
Comparative cognition
False beliefs
Social cognition

ABSTRACT

Human adults flexibly reason about others' unobservable mental states, a capacity known as Theory of Mind (ToM). Unfortunately, the roots of this capacity still remain poorly understood as both infants and nonhuman primates perform inconsistently on tests of ToM. Here, we try to better understand this complex and often contradictory literature by dissecting nonhuman primates' failures on false belief tasks. Across four studies (n = 566 subjects), we find that—despite succeeding on a closely matched control—rhesus monkeys fail to predict how agents with false beliefs will behave even when those agents perform unexpected actions. We interpret this pattern of performance as evidence that monkeys fail to maintain a representation of another agent's recent knowledge once the environment changes outside of that agent's awareness thus preventing monkeys from making any predictions about the agent's future actions. Overall, this work helps to move beyond the success/failure dichotomy typically used to assess ToM performance, and instead provides a more precise characterization of primates' signature limits in ToM, which we speculate may also be shared with human infants.

Imagine that you are waiting for the barista at a coffee shop to make you a cappuccino. As soon as the barista puts your drink on the counter, a stranger takes your drink and then knocks it over when you confront them. Depending on what you observe about the stranger's actions, you might infer that they had a false belief about whose beverage it was and then accidentally knocked it over when you addressed them. Alternatively, you could infer that the stranger knew that the drink was yours and then purposefully knocked it over in anger when they were caught. This capacity to draw such rich and meaningful inferences about people's mental states from their behaviors is what psychologists refer to as *Theory of Mind* (hereafter ToM; Premack & Woodruff, 1978; Wimmer & Perner, 1983).

The central role that ToM plays in humans' social experience raises many questions about the ontogenetic and phylogenetic roots of this capacity. When are children able to represent others' mental states and what earlier, more primitive stages precede a mature ToM? Which ToM-like abilities are shared with our nonhuman primate relatives and which aspects are unique to humans? Developmental and comparative psychologists have sought answers to these questions over the last few decades, and this work has generated complicated and at times contradictory findings.

Consider, for example, one of the most well-known tests of ToM: the Sally-Anne false belief task (Baron-Cohen et al., 1985). In this task, one character, Sally, places a marble in a basket and leaves the scene, and then a second character, Anne, moves the marble from the basket into a box. When three-year-old children are asked where Sally will search for her marble, they fail to predict that Sally will look for her marble in the basket (where she incorrectly believes it to be). Instead, three-year-olds predict that Sally will search for her marble in its current location. Indeed, young children fail to verbally predict how others' false beliefs will shape their behavior in many studies—including those specifically designed to reduce task demands—incorrectly predicting that agents with false beliefs will behave in the same way that a knowledgeable agent would (Shahaeian et al., 2011; Wellman et al., 2001; Wellman & Liu, 2004; Wellman & Woolley, 1990; Wimmer & Perner, 1983).

Interestingly, when two-year-olds are tested using adapted versions of this false belief task, they pre-emptively look to where the agent falsely believes the marble is located (as if correctly anticipating that agent's actions; Garnham & Perner, 2001; Southgate et al., 2007; Surian & Geraci, 2012) and 15-month-old infants look longer when an agent does not act consistently with her false beliefs (indicating an unexpected outcome; Onishi & Baillargeon, 2005; Scott & Baillargeon, 2009). Thus,

* Corresponding author at: 100 College Street, New Haven, CT 06510.
E-mail address: amanda.royka@yale.edu (A. Royka).
[1] Denotes equal contribution.

https://doi.org/10.1016/j.cognition.2025.106400
Received 24 April 2025; Received in revised form 20 November 2025; Accepted 5 December 2025
Available online 29 December 2025

when tested using indirect measures, such as looking time and anticipatory looking, infants seem to be more precocious in their false belief reasoning than previously thought, even though older children consistently fail verbal tasks that test for the same representational capacities.

While researchers have developed numerous theories to accommodate findings like these (e.g., Apperly & Butterfill, 2009; Baillargeon et al., 2010; Carruthers, 2013), such accounts have been complicated by replication attempts that have failed to find consistent evidence for infants' early false belief understanding. Several replication attempts – including direct replications with increased sample sizes – have found that infants and young children do not anticipate the false-belief-driven actions of others (Barone et al., 2022; Dörrenberg et al., 2018; Grosse Wiesmann et al., 2018; Kulke et al., 2018; Kulke & Rakoczy, 2018) and fail to show longer looking when an agent with a false beliefs searches for an object in its true location (Dörrenberg et al., 2018; Kulke & Rakoczy, 2018; Poulin-Dubois & Yott, 2018; Powell et al., 2018). However, some infant ToM results seem more replicable than others. Specifically, researchers have found more consistent evidence that infants have expectations about how knowledgeable agents will behave (Hamlin et al., 2013; Holland & Phillips, 2020; Luo & Baillargeon, 2007; Luo & Johnson, 2009; Vouloumanos et al., 2014; Wellman & Liu, 2004). Yet, researchers to date disagree about the root cause of these failed false belief task replications (Baillargeon et al., 2018; Poulin-Dubois et al., 2018).

Similarly, nonhuman primates' (hereafter, primates) pattern of performance on false belief tasks is also variable. When primates compete against an opponent for food in naturalistic scenarios, they prefer to approach hidden food items that their competitor does not see or know about (Flombaum & Santos, 2005; Hare et al., 2000, 2001; Kaminski et al., 2008; Karg et al., 2015b; Melis et al., 2006), but fail to outcompete competitors with false beliefs (Hare et al., 2001; Kaminski et al., 2008; Krachun et al., 2009, 2010). Primates also perform poorly in looking-time false belief tasks adapted from infant cognition (Drayton & Santos, 2018; Horschler et al., 2019, 2021; Marticorena et al., 2011; Martin & Santos, 2014). In these tasks, primates see a demonstrator witness a desirable food reward (e.g., a lemon) move from one opaque box to a second. The demonstrator's view is then occluded as the lemon moves back from this second box to its original location. Subjects then see the demonstrator reach either into the box where the lemon is hidden or into the empty box. If primates expect an agent with a false belief to search for an object where they believe it to be, then primates should look longer when the demonstrator reaches towards the object's true location. However, primates do not show this pattern of looking and instead look equally long at either outcome. In contrast, recent studies suggest that primates will preemptively look at the location where an agent falsely believes a hidden object is hidden (Hayashi et al., 2020; Kano et al., 2019; Krupenye et al., 2016). Thus, primates could have some understanding of false beliefs that affords successful anticipatory looking performance, but not successful performance in other tasks.

As this complicated pattern of performance demonstrates, the developmental and evolutionary origins of ToM are still very much in question. Before the age of four, young children typically make egocentric errors on verbal ToM tasks: they expect that an agent with a false belief will behave as though she has perfect knowledge of the world (Shahaeian et al., 2011; Wellman et al., 2001; Wellman & Liu, 2004; Wellman & Woolley, 1990; Wimmer & Perner, 1983, but see Fabricius et al., 2021 for a review of errors extending later into childhood). However, the contentiousness of the infant ToM literature makes it hard to draw conclusions about what developmental stages precede this notable error. Similarly, primates consistently fail competition-based and violation of expectation tests of false belief understanding (Drayton & Santos, 2018; Hare et al., 2001; Horschler et al., 2019, 2021; Kaminski et al., 2008; Krachun et al., 2009, 2010; Marticorena et al., 2011; Martin & Santos, 2014), but perform better when tested using anticipatory looking measures (e.g., Krupenye et al., 2016). Such contradictory evidence raises a puzzle both for our understanding of the foundations of ToM and also for the field as a whole. How can developmental and comparative psychologists move forward and build theories that can accommodate all of these nuanced findings? Here, we argue that one path forward involves focusing not on infants' and primates' successes, but on their *failures*. Indeed, cognitive scientists have recently called for a renewed focus on the limits of cognitive capacities across species (Howard & Barron, 2024; Taylor et al., 2022) and this approach has yielded valuable insights into a variety of capacities ranging from transitive inference (Bond et al., 2003; Bryant & Trabasso, 1971; Davis, 1992; McGonigle & Chalmers, 1984; Treichler & Van Tilburg, 1996; von Fersen et al., 1991) to numerical cognition (Agrillo et al., 2011; Barnard et al., 2013; Cantlon & Brannon, 2007; Cordes & Brannon, 2009; Feigenson & Carey, 2003, 2005; Lewis et al., 2005; MaBouDi et al., 2021; Rugani et al., 2010; vanMarle et al., 2006; Xu & Spelke, 2000).

Here, we argue that researchers can gain similar traction on the roots of human ToM by exploring the kinds of errors that infants and primates make on ToM tasks. Importantly, there are striking similarities in the ToM tasks that primates, infants, and young children struggle with (namely, tasks that require them to represent *reality-inconsistent* mental states like false beliefs and ignorance), but not in *the kinds of errors they make*. Consider, for example, the specific kind of error that young children make on verbal false belief tasks: children consistently predict that an agent with a false belief will search for the object in its actual location as if they share the child's knowledge. Interestingly, this type of error is not restricted to agents with false beliefs. Young children also over-attribute knowledge to ignorant agents (Birch & Bloom, 2003; Hogrefe et al., 1986; Mossler et al., 1976; Sullivan & Winner, 1991; Wellman & Liu, 2004), suggesting that children may struggle to represent reality-inconsistent mental states (such as ignorance and false beliefs) in part because those mental states involve content that is inconsistent with their own firsthand knowledge of the world (Phillips et al., 2021; Phillips & Norby, 2021). Note, however, that this pattern of performance has not been observed in infants. Although infants' looking behavior sometimes suggests that they can correctly anticipate the actions of those with false beliefs and other times reflects chance performance, no study of preverbal infant ToM to our knowledge has uncovered a tendency for infants to over-attribute knowledge to agents with reality-inconsistent mental states.

Primates also struggle to predict how agents with reality-inconsistent mental states (i.e., ignorance and false beliefs) will behave, but their errors also look importantly different from those of young verbal children. In contrast to children, primates do not over-attribute their own knowledge when predicting how an agent with a false belief will behave. Monkeys typically perform at chance on false belief tasks (Drayton & Santos, 2018; Horschler et al., 2019, 2021; Marticorena et al., 2011; Martin & Santos, 2014). Take, for example, rhesus macaques' performance on a standard violation of expectation false belief task (e.g., Marticorena et al., 2011). If they had the same expectations as 3-year-old children, then subjects should expect that an agent with a false belief about the location of an object will search for the object in its true location. This expectation should induce rhesus macaques to look longer when the demonstrator reaches to the empty location. But this is not how monkeys perform. Instead, monkeys look equally long when an agent searches for a hidden object in its true location and when that agent searches for the object in the location where the agent should believe it to be. Monkeys, however, perform differently when an agent in this same task has a true belief: when an agent witnesses an object being hidden, monkeys expect her to reach for it in the correct location and look longer when she does not, suggesting that they expect a knowledgeable agent to search in the object's true location (Arre et al., 2021; Drayton & Santos, 2018; Horschler et al., 2019; Marticorena et al., 2011). Monkeys continue to show correct, knowledge-based expectations even when the boxes undergo subsequent rotational displacement (Drayton & Santos, 2018) and when the agent's view is briefly blocked (Marticorena et al., 2011). In this way, the same primates who perform

at chance on false belief tasks also demonstrate robust expectations about a knowledgeable agent's behavior on closely matched tasks. This pattern suggests that monkeys' chance performance on false belief tasks stems not from task demands, but from their inability to accurately predict the behavior of agents with reality-inconsistent mental states like ignorance and false beliefs.

Similarly, when nonhuman great apes, such as chimpanzees, know where a reward is hidden, but have to predict the behavior of an agent that does not, they sometimes perform like monkeys, showing chance performance as if they are unsure where the agent will search (Hare et al., 2001; Krachun et al., 2010). Other times, nonhuman great apes' search behavior appears to be more sensitive to the value of the food reward itself, rather than the competitor's mental states (Kaminski et al., 2008). Notably, in false belief tasks when subjects themselves do not know where a reward is hidden and have to rely on the behavior of a demonstrator to find it, chimpanzees behave as though they expect the demonstrator to be competent (i.e., reaching for or marking the true location of the reward; Call & Tomasello, 1999; Krachun et al., 2009) regardless of whether the agent is knowledgeable or has a false belief. Crucially, however, this pattern is still distinct from the over-attribution of knowledge that human children usually exhibit: a chimpanzee will expect the demonstrator to "get it right" only when the chimpanzee itself lacks firsthand knowledge of the reward's location.

Taken together, infants, young children, and primates often struggle to accurately predict the actions of agents with reality-inconsistent mental states, like false beliefs and ignorance. However, the types of errors exhibited by infants and primates look critically different from those of young verbal children. While researchers have debated whether young children's errors are better characterized as an egocentric over-attribution of their own knowledge or a failure to inhibit a reality-bias (Birch & Bloom, 2003; Gopnik, 1993; Wellman et al., 2001), less attention has been paid to infants' and primates' failures on related tasks. Here, we explore two potential theoretical accounts that could explain their poor performance.

One possibility—which we refer to as the *multiple predictions account*—is that infants and primates struggle on false belief tasks because they form multiple predictions about what an agent with a false belief will do next (e.g., the agent may be equally likely to search in a location where they falsely believe an object to be or in a location where the object actually is). Under this view, infants and primates *do* successfully track and represent some information about what the agent has seen in the past, but, in the end, think it is possible that the agent will search in multiple locations. Such multiple predictions could emerge if infants and primates represent the agent's false belief, but experience interference from features of the experimental set-up. For example, if a desired object is still hidden in a location when the subject has to predict how a demonstrator will behave (e.g., Krachun et al., 2010; Marticorena et al., 2011), then they may experience conflict between their own representation of where the object really is and the demonstrator's false belief. Thus, subjects end up with two conflicting predictions about where the agent will search. This account could explain why both monkeys and nonhuman great apes seem to succeed on false belief tasks more often when the object of interest is removed from the scene (Hayashi et al., 2020; Krupenye et al., 2016).

A second possibility—which we refer to as the *no prediction* account—is that infants and primates struggle with false belief tasks because of representational limits that cause them to drop their expectations about how an agent will behave once that agent's mental states are out of step with reality. For example, infants and monkeys may show chance performance on violation of expectation tests of false belief understanding because they have no expectations to be violated. Under this view, infants and monkeys drop their representation of an agent's past or current mental state when the agent does not observe the object being moved. As such, infants and primates have nothing on which to base their subsequent predictions about what that agent might do next, leading them to look equally long when the agent searches in either location. This account could explain why, in active search versions of false belief tasks, nonhuman great apes seem to sometimes rely on heuristics— such as assuming the experimenter is competent (Call & Tomasello, 1999; Krachun et al., 2009) or prioritizing higher value food (Kaminski et al., 2008)— since they may have no prediction about what an agent with a false belief did.

Distinguishing between these explanations for infants' and primates' ToM failures would not only deepen our understanding of what representational capacities characterize the ontogenetic and phylogenetic roots of ToM, but could also help to build better explanations for the contradictory findings in false belief tasks that currently pervade the literature. Here, we take a first step towards this understanding by focusing more specifically on primates' errors.

To distinguish between the multiple prediction and no prediction hypotheses described above, we tested rhesus macaque monkeys (*Macaca mulatta*) on a standard violation of expectation false belief task, but with the addition of a third irrelevant hiding location. We reasoned that if monkeys generate *multiple predictions* about how an agent with a false belief will behave, then they should find an agent searching in this irrelevant location unexpected given that reaching to this third location should fall outside of the range of predicted actions. On the other hand, if monkeys have *no prediction* about how an agent with a false belief will behave, then witnessing an agent search in a completely irrelevant hiding location should be unsurprising since they have no expectations to be violated.

1. Experiment 1

1.1. Methods

1.1.1. Subjects

We successfully tested 148 rhesus macaque monkeys at the Cayo Santiago Field Station in Puerto Rico (60 females; mean age = 5.46 ± 3.05; see SI for demographics broken down by condition for all studies). This island is home to over 1500 free-ranging and individually identifiable monkeys that are habituated to humans and to participation in violation of expectation paradigms. Additional monkeys were approached but did not complete participation in the study because they were inattentive during critical parts of the study ($n = 16$), left the study area (n = 16), were displaced by another monkey ($n = 15$), had previously participated in part of the study ($n = 28$), or due to an apparatus failure (n = 1). Decisions not to include a subject in the final analysis for this and all subsequent studies were made by the cameraperson who was blind to condition information.

1.1.2. Apparatus

We designed our false belief task to closely parallel ones used previously with this population (Horschler et al., 2019; Marticorena et al., 2011). Monkeys watched a series of trials in which an experimenter interacted with an apple on a small foamcore stage (36 in long, 5.5 in tall, and 6.5 in deep. The stage had a back (14 in high) and three small boxes (4 × 4 × 4 in) equally spaced along the stage. One of the boxes (hereafter the irrelevant box) was separated from the other two by a wall that was connected to the back of the apparatus. We used a small slit in the stage between the two other boxes (hereafter the relevant boxes) to slide an apple in between and into the two boxes surreptitiously. A front occluder (36 × 10.5 in) ran the length of the apparatus, allowing the experimenter to completely hide the stage from the monkeys' view. A smaller top occluder (21.5 × 8.5 in), was attached to the back of the apparatus which could block the experimenter's view (but not the subject's view) of the relevant boxes. The same apparatus was used in all subsequent studies.

1.1.3. Procedure

As in previous studies (e.g., Drayton & Santos, 2017; Marticorena et al., 2011; Rosati & Santos, 2016), experimenters opportunistically

approached monkeys that were sitting calmly in a low-distraction environment. One experimenter (the demonstrator) knelt behind the apparatus approximately 1–2 m away from the subject. A cameraperson stood directly behind the demonstrator to record the looking time of the monkey. Each subject saw two familiarization trials and one test trial.

The *first familiarization trial* was meant to engage the subject and familiarize them with the stage and the apple. The front occluder was flipped down to reveal an apple on the stage in between the relevant boxes (Fig. 1a). The demonstrator said "now" and stared down at the apple for 10 s while the subject's looking was recorded. After the trial, the stage was once again occluded. The cameraperson then gave the demonstrator a letter code that corresponded to a condition (the cameraperson was blind to which codes corresponded to which conditions and thus could stay blind to condition) and dictated the demonstrator's next actions. In the *second familiarization trial*, the demonstrator reached to one of the three boxes (either the irrelevant box or one of the two relevant boxes) based on the condition, said "now" (Fig. 1b) and then paused for 10 s as the monkeys' looking was recorded. The occluder was once again then flipped up to hide the stage. The design of these familiarizations mirrored those used in past ToM studies conducted with this population (Drayton & Santos, 2018; Horschler et al., 2019, 2021).

All subjects then saw a single *test trial*, which began with a belief induction event to establish where the demonstrator thought that the apple was. The demonstrator first watched the apple move out of one of the relevant boxes and into the second one. The top occluder was then flipped to block the demonstrator's view of the relevant boxes and the demonstrator stared at the irrelevant box. While the demonstrator's view was blocked, subjects could see the apple moved back into the box that it was originally in (Fig. 1c). In this way, this belief induction event established that the demonstrator had a false belief about the final location of the apple.

Once the apple finished moving, the top occluder flipped back down and the demonstrator reached into the same box that they reached into during the second familiarization trial (Fig. 1d). This led to three different test conditions: (1) the demonstrator reached into the irrelevant box (*irrelevant reach* condition), (2) the demonstrator reached into the box where they last saw the apple (*false-belief-consistent reach* condition), or (3) the demonstrator reached into the box where the apple was actually located (*reality-consistent reach* condition). Once the demonstrator made their reach, he said "now" and held the pose for 10 s while subject looking was recorded.

In this and all subsequent experiments, we counterbalanced which box the apple started in across subjects. Additionally, at the halfway point of data collection, we switched the location of the top occluder and the wall separating the irrelevant box, such that the box on the opposite end of the apparatus became the irrelevant box.

1.1.4. Video coding

All trials were independently analyzed by two coders (one using Adobe Premiere Pro 2022 and the second using BORIS). The videos were clipped and randomized such that coders were blind to the condition. Each video was clipped to 10 s, starting immediately after the experimenter said "now." The coders indicated how long the subjects looked at either the apparatus or the experimenter during the 10 s. Inter-rater reliability was high (Pearson's $R = 0.93$).

1.2. Results

We used a linear mixed effects model to predict subjects' log-transformed looking times in a given trial based on the trial type (familiarization 1, familiarization 2, or test), condition (irrelevant reach, false-belief-consistent reach, or reality-consistent reach) and their interaction with random intercepts for each subject. For this and all other regressions reported, each regression was rotated in order to change the baseline levels of the dummy-coded trial type and condition variables (see SI for full rotated regression tables). Across all three conditions, subjects' looking times significantly decreased from familiarization 1 to familiarization 2 ($\beta = -0.25, p = .008$ using the irrelevant condition as the baseline) with no significant interaction between trial type and condition (for looking time changes from familiarization 1 to familiarization 2 comparing the irrelevant reach to false-belief-consistent reach: $\beta = -0.21$, $p = .115$; comparing irrelevant reach to reality-consistent reach: $\beta = -0.08$, $p = .537$; comparing false-belief-consistent reach to reality-consistent reach: $\beta = 0.13$, $p = .325$; see Fig. 2). This pattern of performance suggests that subjects were familiarized as expected before the test events.

Moreover, there were no significant differences in subject looking times at test across the three conditions (comparing irrelevant reach to false-belief-consistent reach: $\beta = 0.02$, $p = .848$; comparing irrelevant reach to reality-consistent reach: $\beta = 0.08$, $p = .484$; comparing false-belief-consistent reach to reality-consistent reach: $\beta = 0.06$, $p = .617$). These results suggest that although subjects were successfully familiarized to the set-up across the three conditions, they did not find any of the reaches in the test trial to be more unexpected than the others.

1.3. Discussion

Although monkeys across all three conditions were successfully familiarized to our looking time set-up, subjects showed no differences in looking across conditions on the test trial. This pattern of results replicates a consistent finding in comparative ToM studies: monkeys look equally long no matter how an agent with a false belief behaves (Drayton & Santos, 2018; Horschler et al., 2019; Marticorena et al.,

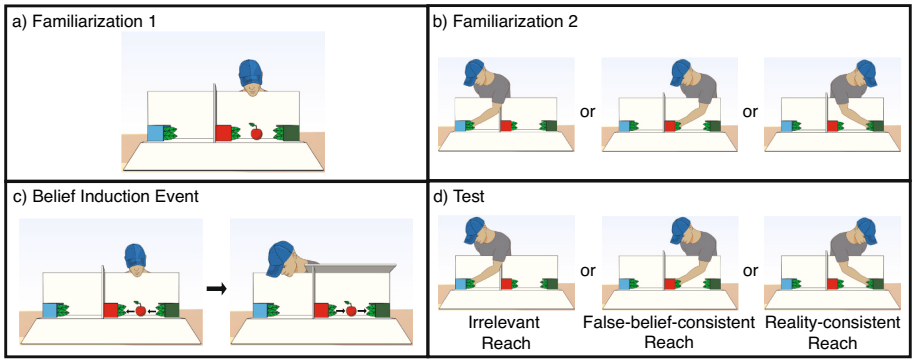

Fig. 1. The procedure for Study 1. a) Subjects first saw the demonstrator staring at the apple in Familiarization 1. b) Then, in Familiarization 2, subjects saw one of three possible reaches depending on a blindly-assigned condition. c) The belief induction event next established the demonstrator's false belief about the location of the apple. d) In the test event, the demonstrator reached to the same box that they did in Familiarization 2, such that the demonstrator made one of three possible reaches: to the box that never held the apple (irrelevant reach), to the box where they last saw the apple (false-belief-consistent reach), or to the box where the apple actually was (reality-consistent reach).

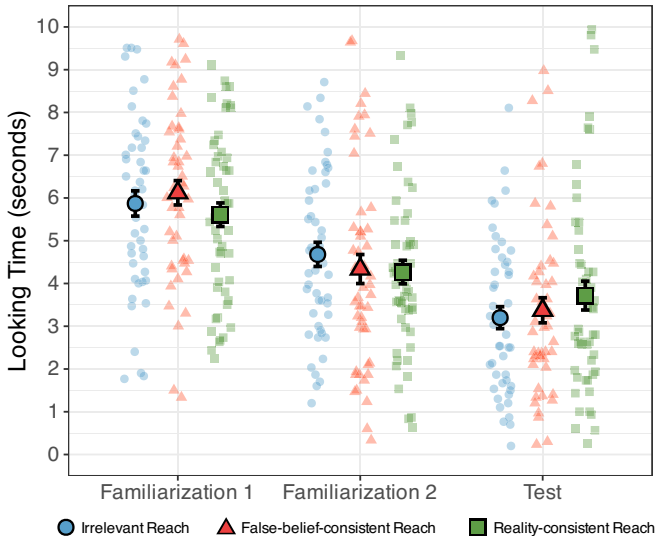

Fig. 2. The results of Study 1 showing subjects' raw looking times (y-axis) across familiarization 1, familiarization 2, and the test trial (x-axis). Dot shape and color indicate the type of reach seen by subjects in the test trial and darker colored shapes represent condition averages with standard error bars.

2011). However, this study goes beyond previous work to show that monkeys also show equal looking when an agent reaches to a completely irrelevant box, one that never held the object and was physically separated from the other boxes by a barrier.

Experiment 1's results are consistent with the no prediction account of primate false belief performance. Subjects' looking responses suggest that no expectation was violated when the demonstrator reached to the irrelevant box. Note that if subjects had had multiple predictions about how the agent would behave (possibly based on the demonstrator's past visual access and the true location of the apple), then they should have looked longer at the irrelevant reach relative to the other two conditions.

One alternative interpretation of Experiment 1's findings, however, is that subjects may have formed multiple predictions but those predictions included the possibility that the demonstrator could reach to the irrelevant box. In an effort to emphasize that the demonstrator was aware that nothing could enter the irrelevant box during the belief induction event, the demonstrator stared at the irrelevant box while the apple moved between the two relevant boxes. Unfortunately, this feature of our design could have had the opposite effect, causing subjects to predict that the agent might reach into the irrelevant box since he showed interest in that location. If staring at a location is sufficient to lead subjects to form an expectation that an agent will reach there, then monkeys in Experiment 1 could have had positive predictions that the agent would reach towards any of the three boxes. We attempted to rule out this alternative in Experiment 2.

2. Experiment 2

Experiment 2 directly tested whether staring at the irrelevant box leads subjects to predict that the agent will reach to that location. We reasoned that if the demonstrator is completely ignorant about which of the two relevant boxes the apple ends up in, but stares at the irrelevant box, then subjects may make a positive prediction that he will search in that box and may also predict that he will search in the box where the object actually is (due to interference from their firsthand knowledge). Crucially, however, subjects should have no such prediction for the other empty relevant box. Thus, when the demonstrator reaches for the relevant, but empty box, the monkeys should find this event unexpected and look longer. In contrast, if the no prediction account is correct, monkeys should look equally long regardless of matter where the demonstrator reaches.

2.1. Methods

2.1.1. Subjects

We successfully tested 147[2] rhesus macaque monkeys (61 females; mean age = 5.23 ± 3.32). Additional monkeys were tested, but excluded because they were inattentive during critical parts of the study ($n = 16$), left the study area ($n = 30$), were displaced by another monkey ($n = 5$), had previously participated in part of the study ($n = 17$), or experimenter error ($n = 4$).

2.1.2. Procedure

As in Experiment 1, all subjects saw two familiarization trials and one test trial. The first and second familiarization trials were identical to those in Experiment 1. Subjects' looking time was recorded for 10 s after each familiarization.

The test trial began with a belief induction event, which established that the demonstrator was ignorant about the location of the apple (rather than having a false belief, as in Experiment 1). As soon as the front occluder dropped revealing the stage to the subject, the top occluder was flipped to block the demonstrator's view of the relevant boxes. While the demonstrator's view was blocked, he stared at the irrelevant box. The subject, but not the demonstrator could see the apple move out of one of the relevant boxes into the second one, and then back into the box where it originally was (Fig. 3a). Therefore, the demonstrator had no belief about which of the two relevant boxes held the

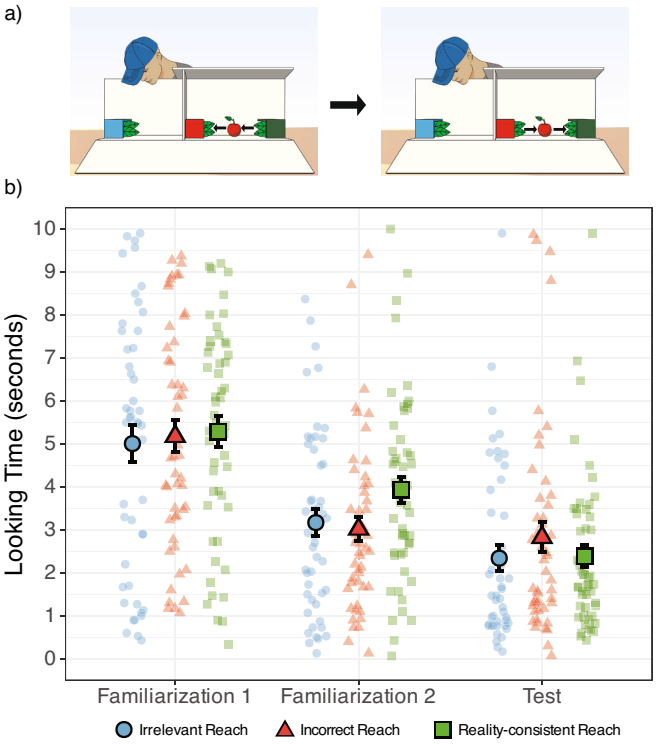

Fig. 3. a) The belief induction event in Experiment 2 established that the demonstrator was ignorant about the exact location of the apple. b) The results of Experiment 2 with subjects' raw looking times on the y-axis across familiarization 1, familiarization 2, and the test trial (x-axis). Dot colors indicate the type of reach seen by subjects in the test trial and darker colored diamonds represent condition averages with standard error bars.

[2] Subjects were allowed to participate in more than one study. Across the four studies presented here, 117 subjects participated in multiple studies with an average of 230.94 days in between studies.

apple.

Once the apple finished moving, the top occluder flipped back down and the demonstrator reached into the same box that he reached into during the second familiarization trial. Thus, the demonstrator either reached into the irrelevant box (*irrelevant reach* condition), the box where the apple was not located (*incorrect reach* condition), or the box where the apple was actually located (*reality-consistent reach* condition). Once the demonstrator made his reach, he said "now" and held the pose for 10 s while subject looking was recorded.

2.1.3. Video coding

All trials were independently analyzed by two coders blind to condition (using Adobe Premiere Pro 2022) with high inter-rater reliability (Pearson's $R = 0.91$).

2.2. Results

Using a linear mixed effects model to predict subjects' log-transformed looking times based on the trial type (familiarization 1, familiarization 2, or test), condition (irrelevant reach, false-belief-consistent reach, or reality-consistent reach) and their interaction (with random intercepts for each subject), we find a significant effect of trial type on looking-time. Specifically, subjects' looking time significantly decreased from familiarization 1 to familiarization 2 ($\beta = -0.49$, $p < .001$). There was no significant interaction between trial type and condition (for looking time changes from familiarization 1 to familiarization 2 comparing irrelevant reach to incorrect reach: $\beta = -0.13$, $p = .489$; comparing irrelevant reach to reality-consistent reach: $\beta = 0.15$, $p = .421$; comparing incorrect reach to reality-consistent reach: $\beta = 0.27$, $p = .132$). As in Study 1, subjects' looking times during the test trial were not significantly predicted by condition (comparing irrelevant reach to incorrect reach: $\beta = 0.20$, $p = .226$; comparing irrelevant reach to reality-consistent reach: $\beta = 0.14$, $p = .396$; comparing false-belief-consistent reach to reality-consistent reach: $\beta = -0.06$, $p = .710$; Fig. 3b).

2.3. Discussion

Subjects in all conditions showed significant decreases in looking time across the familiarization trials, suggesting that they were sufficiently familiarized with the set-up. Nevertheless, subjects still showed no difference in looking across the three test conditions. Even though the demonstrator in Experiment 2 stared at the irrelevant box for an even longer duration than in Experiment 1, subjects were not surprised when the demonstrator reached to either of the other two boxes. This pattern of results suggests that simply seeing an agent look at a location is not sufficient to cause monkeys to form an expectation that the agent will subsequently act on that location. Thus, the results of Experiment 1 cannot be explained by positing that monkeys formed multiple predictions about where the demonstrator would act.

Experiment 2's findings are once again consistent with the no prediction account. As soon as the demonstrator's view is occluded and the apple begins moving during the test event, the mismatch between the demonstrator's perspective and the monkey's own perspective seems to lead monkeys to form no predictions about how an agent will behave regardless of whether that agent has a false belief (as in Experiment 1) or is ignorant (as in Experiment 2).

One alternative explanation, however, is that monkeys represented the demonstrator as ignorant and therefore formed two positive expectations that he could reach into either relevant box in addition to a third expectation that he might reach into the irrelevant box since he was staring at it. Past work suggests that rhesus macaques do not rely on these sorts of ignorance representations in similar VOE tasks (for a review, see Horschler et al., 2019, 2021; Martin & Santos, 2016). However, to provide further evidence in favor of the no prediction account, our next experiment prevented the demonstrator from staring at the irrelevant box as the apple changed location.

Experiment 3 also increased the irrelevance of the third box by showing both the subject and the demonstrator that it was empty. Experiments 1 and 2 support the no prediction account of primates' false belief performance, as primates seemed to make no prediction about where an agent with a reality incongruent mental state would reach for a desirable object. This general pattern is consistent with the proposal that monkeys, under certain circumstances (see Horschler et al., 2020; Martin & Santos, 2016), drop representations of others' mental states and fail to form expectations about how they will behave. However, the no prediction account would also argue that primates will still perform at chance even if the irrelevance of one box is substantially increased. Here, we test this prediction by not only having the third irrelevant box physically separated from the other two, but also by explicitly showing subjects that the demonstrator knows that this irrelevant third location is empty.

3. Experiment 3

3.1. Methods

3.1.1. Subjects

121 monkeys participated (54 females; mean age = 6.31 ± 4.31). Additional monkeys were excluded: 67 for inattention, 48 for leaving the study area, 8 for displacement by another monkey, 23 for previous participation, 2 for experimental errors, and 6 for approaching the apparatus.

3.1.2. Procedure

All subjects saw the same first and second familiarization trials used in Experiments 1 and 2 (Fig. 1a-b). However, the belief induction event was changed to heighten the irrelevance of the third box. In this new version of the belief induction event, the demonstrator first watched the irrelevant box flip open, revealing that it was empty, and then flip closed again (Fig. 4a). The demonstrator then watched the apple move out of one of the relevant boxes and into the second one. The top occluder then flipped to block the demonstrator's view of the relevant boxes, but this time stayed upright, parallel to the back of the stage, blocking the demonstrator entirely from the subject's view (Fig. 4a). While the demonstrator's view was blocked, subjects could see the apple move back into the box that it was originally in (Fig. 4a). Thus, the demonstrator knew the irrelevant box was empty and had a false belief about the location of the apple, but did not look at the irrelevant location as the apple switched locations.

Once the apple finished moving, the top occluder flipped back down and the demonstrator reached into the same box that he reached into during the second familiarization trial. As in Experiment 1, this led to three different test conditions: (1) the demonstrator reached into the highly irrelevant box (irrelevant reach condition), (2) the demonstrator reached into the box where they last saw the apple (false-belief-consistent reach condition), or (3) the demonstrator reached into the box where the apple was actually located (reality-consistent reach condition).

3.1.3. Video coding

Coding was the same as in Experiment 1 with high inter-rater reliability (Pearson's $R = 0.94$).

3.2. Results

As in Experiments 1 and 2, we used a linear mixed effects model to predict log-transformed looking times in a given trial based on the trial type, condition and their interaction with random intercepts for each subject. Although subjects in the reality-consistent reach condition had looking times that significantly decreased between familiarization 1 and familiarization 2 ($\beta = -0.88$, $p < .001$), there was a significant

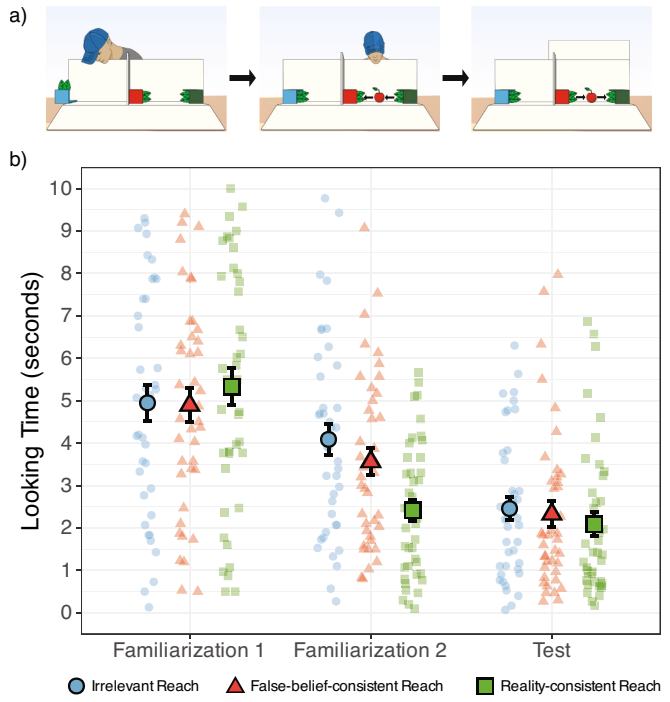

Fig. 4. a) During Experiment 3's belief induction event, the demonstrator first saw that the irrelevant box was empty, then observed the first location change, but was out of sight for the final location change. Thus, Experiment 3's belief induction event increased the irrelevance of the third box and established that the demonstrator had a false belief about the location of the apple. b) The results of Experiment 3 with subjects' raw looking times on the y-axis across familiarization 1, familiarization 2, and the test trial (x-axis). Dot colors indicate the type of reach seen by subjects in the test trial and darker colored diamonds represent condition averages with standard error bars.

interaction such that subjects in the other two conditions had looking times that decreased significantly less compared to subjects in the reality-consistent reach condition (for looking time changes from familiarization 1 to familiarization 2 comparing the reality-consistent reach to the irrelevant reach: $\beta = 0.72$, $p < .001$; comparing reality-consistent reach to the false-belief-consistent reach: $\beta = 0.58$, $p = .004$). Since subjects across all conditions saw the same familiarizations, this interaction is difficult to interpret, but could indicate that, by chance, monkeys in the reality-consistent reach condition habituated to the set-up more quickly. Critically, however, across all three conditions, subjects' looking times were significantly lower at test compared to familiarization 1 ($\beta = -0.78$, $p < .001$ using the irrelevant condition as the baseline) with no significant interaction between trial type and condition (for looking time changes from familiarization 1 to test comparing the irrelevant reach to false-belief-consistent reach: $\beta = -0.21$, $p = .115$; comparing irrelevant reach to reality-consistent reach: $\beta = -0.09$, $p = .653$; comparing irrelevant reach to reality-consistent reach: $\beta = -0.28$, $p = .157$). As such, we can be confident that monkeys in this study were not attending at ceiling levels throughout the entire presentation and that their looking times could have increase at test in response to an unexpected event.

At test, however, there were no significant differences in subject looking times at test across the three conditions (comparing irrelevant reach to false-belief-consistent reach: $\beta = -0.05$, $p = .797$; comparing irrelevant reach to reality-consistent reach: $\beta = -0.19$, $p = .308$; comparing false-belief-consistent reach to reality-consistent reach: $\beta = -0.14$, $p = .447$; Fig. 4b). Thus, subjects still showed no violation of expectation response when the demonstrator reached to a truly irrelevant empty box.

3.3. Discussion

Experiment 3's results were once again consistent with the no prediction account; subjects in Experiment 3 also seemed to generate no predictions about how the demonstrator would behave. Crucially, subjects in Experiment 3 showed no prediction even when the demonstrator reached to a box that he saw was empty less than a minute earlier. Thus, subjects persisted in showing no violation of expectancy response even when faced with an agent performing an irrational action that contradicted his prior knowledge. This somewhat unusual result is fully consistent with the no prediction account, which argues that primates drop *all representations* of an agent's mental states when there is a mismatch between the demonstrator's perspective and the primate's own perspective. Additionally, the results of Experiment 3 are also inconsistent with a multiple predictions account of primates' performance; there was no reason why primates tested in Experiment 3 should form a prediction that reaching into the fully irrelevant box would be just as likely as reaching into the two more relevant boxes.

4. Experiment 4

Although the results of Experiments 1 through 3 provide support for the no prediction account of primates' false belief performance, this evidence so far has only come in the form of primates' chance performance. One could therefore interpret primates' performance thus far not as evidence for a no prediction account but instead as chance performance based on the task being too complicated overall. As such, Experiment 4 aimed to test whether primates are capable of making successful predictions about an agent's behavior in the same set-up when the agent has a reality-*consistent* mental state— namely when he is fully aware of where the object is located. Previous work has shown that primates are able to correctly predict where a knowledgeable agent will act both in novel and naturalistic set-ups (Arre et al., 2021; Drayton & Santos, 2017, 2018; Flombaum & Santos, 2005; Hare et al., 2000, 2001; Horschler et al., 2019; Kaminski et al., 2008; Karg et al., 2015b; Krachun et al., 2009; MacLean & Hare, 2012; Marticorena et al., 2011; Melis et al., 2006; Santos et al., 2006). If subjects in Experiments 1–3 showed chance performance because monkeys were confused by our complicated apparatus, then they should also fail at making predictions on the same set-up when the agent knows where the object is. In contrast, if subjects performed at chance in Experiments 1–3 specifically because they are unable to make predictions about the behavior of agents with reality-inconsistent mental states like ignorance and false beliefs, then subjects in Experiment 4 should succeed in predicting how an agent with a reality-consistent mental state should behave and thus they should look longer when a knowledgeable agent reaches to one of the incorrect two boxes than when he reaches to correct box.

4.1. Methods

4.1.1. Subjects

We tested 150 monkeys (72 females; mean age = 5.20 ± 3.28) for Experiment 4. Additional monkeys were approached but did not complete participation due to inattention ($n = 14$), leaving the area ($n = 33$), displacement ($n = 12$), previous participation ($n = 41$), approach ($n = 2$), or due to an experimenter error (n = 2).

4.1.2. Procedure

The familiarization trials and condition assignment were identical to Experiment 1. However, during the test trial, the demonstrator's view was not occluded. Instead, he watched the apple move out of one of the relevant boxes, into the second one, and then return to its original location (Fig. 5a). Thus, the demonstrator witnessed all of the apple's movements, making him knowledgeable about the apple's final location.

Once the apple finished moving, the demonstrator reached into the

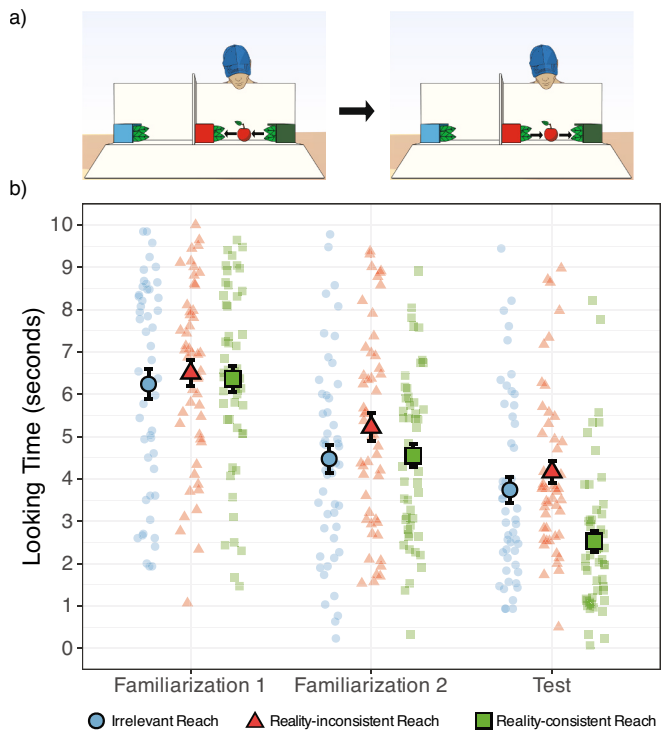

Fig. 5. a) Experiment 4's belief induction event established that the demonstrator had a trued belief about the location of the apple. b) Experiment 4 results. Crucially, subjects' looking times (y-axis) decreased significantly more in the reality-consistent reach condition, suggesting that subjects found this outcome less surprising that a reach to the irrelevant and to the reality-inconsistent box. Darker colored diamonds represent condition averages with standard error bars.

same box that he reached into during the second familiarization trial. This led to three different test conditions: (1) the demonstrator reached into the irrelevant box (*irrelevant reach* condition), (2) the demonstrator reached into the box where the apple was not located (*knowledge-inconsistent reach* condition), or (3) the demonstrator reached into the box where the apple was located (*knowledge-consistent reach* condition). Once the demonstrator made his final reach, he said "now" and held the pose for 10 s while subject looking was recorded.

4.1.3. Video coding

Trials were coded as in previous studies with high inter-rater reliability (Pearson's $R = 0.92$).

4.2. Results

Using a linear mixed effects model predicting subjects' log-transformed looking times in a given trial based on the trial type (familiarization 1, familiarization 2, or test), condition (irrelevant reach, knowledge-inconsistent reach, or knowledge-consistent reach) and their interaction with random intercepts for each subject, we found that trial type significantly predicted subjects' looking times: looking times were significantly lower in familiarization 2 relative to familiarization 1 ($\beta = -0.38$, $p < .001$ using the reality-consistent reach condition as the baseline) with no significant interaction between trial type and condition (for looking time changes from familiarization 1 to familiarization 2 comparing the reality-consistent reach to the irrelevant reach: $\beta = -0.05$, $p = .750$; comparing reality-consistent to reality-inconsistent reach: $\beta = 0.12$, $p = .413$; comparing irrelevant reach to reality-inconsistent reach: $\beta = 0.16$, $p = .256$). This pattern of performance suggests that subjects were familiarized as expected before the test events.

In the test trial, subjects looked significantly longer after seeing a reality-inconsistent reach relative to a reality-consistent reach ($\beta = 0.66$, $p < .001$), replicating past studies (e.g., Horschler et al., 2019; Kaminski et al., 2008). Moreover, subjects looked significantly longer after seeing an irrelevant reach relative to a reality-consistent reach ($\beta = 0.49$, $p < .001$). There was no significant difference in test trial looking between the irrelevant and reality-inconsistent reach conditions ($\beta = 0.17$, $p = .144$; Fig. 5b). Thus, monkeys spent less time looking when the demonstrator's actions were consistent with his knowledge. However, when the demonstrator reached to either of the two empty boxes, monkeys looked longer, suggesting that they found these outcomes more unexpected.

4.3. Discussion

In contrast to their performance in Experiments 1–3, monkeys tested in Experiment 4 expected the demonstrator to behave in line with his mental state. Specifically, monkeys showed a significant increase in looking when a knowledgeable agent behaved in a manner that did not align with his knowledge (i.e., reaching towards the irrelevant box or the box where the apple was no longer located), suggesting that monkeys found these events to be unexpected. Thus, monkeys can make a positive prediction about whether or not an agent should reach to an irrelevant location in our experimental set-up, but they're only able to do so when that agent has accurate knowledge of the object's location.

The results of Experiment 4 also provide evidence against possible alternative explanations for subjects' null performance in our earlier studies. For example, the larger size of our apparatus relative to previous studies with this population (e.g., Drayton & Santos, 2018; Horschler et al., 2019, 2021; Marticorena et al., 2011) could have caused subjects to disengage from the task possibly due to stress or difficulty tracking the possible hiding locations. However, we found that subjects succeeded in making predictions about the demonstrator's search behavior towards the irrelevant box in Experiment 4, suggesting that subjects' earlier performance was not undermined by the design of our apparatus. A different account for the failures that we observed in Experiments 1–3 is that subjects' performance could have been affected by the specific event they saw in familiarization 2 (in which the demonstrator reaches to the location where they later reach in the test trial). More specifically, it's possible that familiarization 2 set up an expectation that the demonstrator would always search in the location where they previously reached. Critically, though, subjects in Experiment 4 showed significantly longer looking in the irrelevant and reality-inconsistent conditions in spite of the fact that they had previously seen the demonstrator reach to those exact locations.

One further alternative account for the null findings in Experiments 1–3 is that monkeys can only make predictions about an agent's behavior based on the presence of an object, but they cannot predict how an agent will behave towards an empty location. If this account were true, then it would be possible that monkeys could represent others' false beliefs, but still fail Experiments 1–3 which require them to track an empty box. Under this account, monkeys would be incapable of predicting how an agent would behave towards the empty irrelevant location and thus end up with no prediction – not due to limits on their ToM representational capacities, but perhaps due to difficulties related to representations of absence. This alternative explanation, however, is also incompatible with the results of Experiment 4. Subjects in the irrelevant reach condition looked significantly longer than subjects in the reality-consistent reach condition, this suggests that subjects did not lack a prediction about how the demonstrator should behave in relation to the empty box. Instead, this pattern suggests that monkeys expected the demonstrator to reach where he knew the apple was and not to either of the other two locations. If subjects were unable to form predictions about how agents behave in relation to empty locations, then Experiment 4 should have also yielded null results across all three conditions since the lack of any prediction would not result in the longer looking

times seen here as there would have been no expectations to be violated.

Altogether, the results of Experiment 4 suggest that the null performance in Experiments 1–3 was not due to the specifics of our previous experimental designs or representational difficulties related to empty locations, but instead stemmed from the fact that monkeys make no prediction when an agent has a reality-inconsistent mental state.

5. General discussion

In addition to conceptually replicating past studies in which monkeys have shown null performance on false belief tasks (Drayton & Santos, 2018; Horschler et al., 2019; Marticorena et al., 2011), the present findings reveal the full extent of monkeys' failures, and in doing so begin to shed new light on why those failures may emerge. Specifically, we find that monkeys show no increase in looking when an agent with a false belief reaches to a completely irrelevant location, suggesting that monkeys do not find this irrational action to be unexpected. We found this pattern not only when the irrelevant location was physically separated from the other possible locations (Experiments 1–3), but also when the agent explicitly saw that the irrelevant location was empty (Experiment 3). Crucially, Experiment 4 found that monkeys can make positive predictions about how an agent with a true belief will behave in a closely-matched set-up: when the agent knew about the apple's final hiding location, monkeys *did* find it unexpected when agent subsequently reached to an irrelevant box where the apple was never located. This pattern of failures reveals a key limitation on monkeys' capacity to predict others' actions: when an agent has a false belief, they seem to be unable to generate *any predictions* about how that agent will behave.

Although our results provide further insight into the nature of errors on comparative ToM tasks, they do raise a new question: *why* do monkeys fail to generate predictions in these cases? One possibility is that monkeys are capable of representing others' mental states only when those mental states are consistent with reality (i.e., monkeys can represent what an agent knows only when they themselves know that same thing; Martin & Santos, 2016; Phillips et al., 2021; Phillips & Norby, 2021). Under this view, once an agent's awareness is out of step with the subject's own (for example, when the agent does not see an object switch locations), monkeys no longer represent that agent's mental state since its content has become reality-inconsistent. The present findings not only corroborate this theory, but also illustrate what happens to monkeys' representations of others' mental states once they become inconsistent with reality: namely, such representations are dropped completely and fail to inform any subsequent predictions about the agent's future behavior.

In this way, primates may exhibit an interesting signature cognitive limit when representing the perspectives of others. Once primates detect that another agent's perspective is out of step with their own, they may drop all representations of that agent's past knowledge and subsequently have no representations on which to base their predictions. In Experiments 1–3, this representational limit led subjects to show equal looking even when the agent irrationally reached to an irrelevant location. Such a signature cognitive limit could also explain why subjects in Experiment 3 failed to predict that the demonstrator should avoid the irrelevant box after seeing that it was empty: the apple's subsequent location change (which was not observed by the demonstrator) caused subjects to drop their representation of the demonstrator's prior knowledge about the contents of the irrelevant box.

There are, however, interesting additional boundaries on this possible signature limit. In Experiments 1–3, subjects seemed to drop their representations of the demonstrator's mental state after he did not see the apple change its location. However, past work suggests that other kinds of unobserved changes *do not* interrupt monkeys' capacity to predict what an agent will do next. For example, if a desired object changes color or if the box containing the object moves and then returns to its original location outside of the demonstrator's view, monkeys still expect the demonstrator to search for the object in its true location (Horschler et al., 2019, 2021). Taken together, these earlier change-of-location ToM tasks and the findings presented here more concretely outline the limitations of rhesus macaques' ToM-like capacities, establishing both what sorts of changes interrupt monkey's capacity to predict what an agent will do next and what sorts of errors should result from those interruptions.

Our studies shed light on the representations that rhesus monkeys use to track the mental lives of others, but do these findings also provide insight into the ToM capacities and limitations of nonhuman great apes? Nonhuman great apes are, of course, more closely related to humans relative to rhesus macaques, and several studies have found that nonhuman great apes do track others' reality-inconsistent mental states to some extent (Buttelmann et al., 2017; Kano et al., 2019; Krupenye et al., 2016; Lurz et al., 2022; Lurz et al., 2025; Townrow & Krupenye, 2025). In spite of these differences between monkeys and apes, we argue that the *no prediction* account discussed here may provide some hints about nonhuman great apes' failures in ToM tasks as well. Consider, for example, some of the specific errors that nonhuman great apes make on ToM tasks. In one study, chimpanzees got frustrated with an experimenter for not giving them a desirable food, even though the experimenter was ignorant of the food's existence (Engelmann et al., 2022). If nonhuman great apes make no predictions about how an ignorant agent will act (and thus do not realize that it is unlikely that an ignorant agent will act on an object they are unaware of), then this could explain the heightened emotional displays. Similarly, a no prediction account also explains the odd performance of chimpanzees tested in false belief tasks in which they themselves do not know where a food reward is hidden. In these tasks, chimpanzees seem to expect an agent with false beliefs to behave competently, and incorrectly follow his reaching behavior or placement of a physical marker (Call & Tomasello, 1999; Krachun et al., 2009). Under a no prediction account, chimpanzees would drop their representations of what the agent believed as soon as he developed a false belief and thus may likely incorrectly follow any available social cues that occurred after that false belief event. Finally, the no prediction account is also consistent with recent findings concerning nonhuman great apes' informative behaviors. Studies have found that chimpanzees and bonobos successfully inform ignorant conspecifics about an unobserved threat (Crockford et al., 2012, 2017) and successfully reveal hidden food to ignorant cooperators who can assist them in accessing it (Karg et al., 2015a; Townrow & Krupenye, 2025). However, great apes tend *not* to hide information that will make knowledgeable agents ignorant. That is, nonhuman great apes fail to actively create situations in which an agent would have a reality-*inconsistent* representation. For example, chimpanzees who are given the opportunity to preemptively hide food from a competitor, fail to do so (Karg et al., 2015a). Taken together, these results are consistent with the idea that nonhuman great apes—like monkeys—may have no predictions about how agents with a false belief will behave.

While the no prediction account can accommodate nonhuman great apes' patterns of errors, how can we rectify this proposal with evidence that apes perform well on some false belief ToM tasks (Buttelmann et al., 2017; Kano et al., 2019; Krupenye et al., 2016; Lurz et al., 2022; Lurz et al., 2025)? We see two possible explanations for how the no prediction account could fit with these successes. First, it is possible that nonhuman great apes can represent others' false beliefs, but that those representations are either fragile or deployed only in certain circumstances (perhaps only in cases of high social drama as suggested by Krupenye et al., 2016). Indeed, recent computational approaches argue that the effect sizes observed in comparative ToM studies hint that primates may only rarely deploy full representational capacities (Berke et al., 2025). This view could explain why nonhuman great apes sometimes succeed on ToM tasks, and why they show similar signature limits as monkeys show here when they fail on those tests. A second possibility is that more critical attention needs to be paid to the strength of positive evidence that nonhuman great apes can represent others' false beliefs. Studies of ape false belief reasoning are sometimes plagued

with only marginal statistical significance (Lurz et al., 2022, 2025). Other studies showing positive evidence for ape false belief reasoning have not always conceptually replicated effects with all dependent measures used (Kano et al., 2019; Krupenye et al., 2016) or have found that apes fail in control conditions that may undermine the interpretability of key results (Buttelmann et al., 2017; see Horschler et al., 2020 for an extended discussion of these issues). Given these limitations, some have suggested that nonhuman great apes' successes on false belief tasks should be interpreted cautiously and we hope that the framework presented here might help to further test the limits of nonhuman great ape ToM.

Crucially, it's important to note that the failures and signature limits we observe here should not be viewed as an attack on the complexity of nonhuman primate social cognition more generally. The present results are consistent with numerous studies showing that primates make robust predictions when agents have reality-*consistent* mental states, such as seeing and knowing. Indeed, a wealth of studies have shown that primates successfully make predictions about the behavior of knowledgeable agents in a variety of contexts (Arre et al., 2021; Drayton & Santos, 2017, 2018; Flombaum & Santos, 2005; Hare et al., 2000, 2001; Horschler et al., 2019; Kaminski et al., 2008; Karg et al., 2015b; Krachun et al., 2009; MacLean & Hare, 2012; Marticorena et al., 2011; Melis et al., 2006; Santos et al., 2006), highlighting the flexibility of this system. In this way, our findings do not undermine the sophistication of primate social cognition, but rather work towards more clearly articulating the representations (and representational limits) that support their behavioral flexibility.

Our findings also open new opportunities for exploring the nature of infants' chance performance on false belief tasks. Researchers have argued that infants perform poorly on some versions of the false belief task because certain task features cause infants' false belief representations to become "contaminated" by their own knowledge of the most recent location of a hidden object (Baillargeon et al., 2018). This proposal is not unlike the multiple predictions account we described earlier. Alternatively, it is possible that infants perform poorly on some false belief tasks because they, like primates, fail to form *any* expectations about the behavior of agents with false beliefs. Under this view, human ToM may also begin in a restricted form, with infants representing and predicting others' behavior only when other agents have reality-consistent mental states. Developmental researchers could distinguish between these two hypotheses by testing human infant participants in the experimental set-up we developed here. The results of such a study would not only provide important insights into the foundations of ToM in humans, but could help to explain possible differences between infants and primates as well.

CRediT authorship contribution statement

Amanda Royka: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Conceptualization. **Daniel J. Horschler:** Writing – review & editing, Project administration, Methodology, Investigation, Conceptualization. **Walker Bargmann:** Writing – review & editing, Investigation. **Laurie R. Santos:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Acknowledgements

The authors would like to thank Nicole Dirks, Sabrina Layton, Madeline Meade, McKenna Ragen, and Rebecca Sosa-Coba for their help in data collection and coding. The authors are also grateful to Alyssa Arre, Giselle Caraballo Cruz, Francelia Martinez, and Nahiri Rivera Barreto for their leadership and support at the Cayo Santiago field station. This work was approved by the Cayo Santiago IACUC committee and conforms to guidelines for the use of animals in research. The Cayo Santiago population is currently supported by the Office of Research Infrastructure Programs (ORIP) of the National Institutes of Health, grant number P40 OD012217. This work was also supported by NSF SBE Postdoctoral Research Fellowship under grant number 2104589 awarded to DJH. The content of the publication is solely the responsibility of the authors and does not necessarily represent the official views of the ORIP or the National Science Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106400.

Data availability

All data, code for analyses, and pre-registrations can be found at https://osf.io/7n6fp/?view_only=4170992a97404002ae7682ce1c33d33d

References

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition, 119*(2), 281–287. https://doi.org/10.1016/j.cognition.2010.10.022

Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970. https://doi.org/10.1037/a0016923

Arre, A. M., Stumph, E., & Santos, L. R. (2021). Macaque species with varying social tolerance show no differences in understanding what other agents perceive. *Animal Cognition, 24*(4), 877–888. https://doi.org/10.1007/s10071-021-01485-5

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development, 46*, 112–124. https://doi.org/10.1016/j.cogdev.2018.06.001

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Barnard, A. M., Hughes, K. D., Gerhardt, R. R., DiVincenti, L., Bovee, J. M., & Cantlon, J. F. (2013). Inherently analog quantity representations in olive baboons (*Papio anubis*). *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00253

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition, 21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Barone, P., Wenzel, L., Proft, M., & Rakoczy, H. (2022). Do young children track other's beliefs, or merely their perceptual access? An interactive, anticipatory measure of early theory of mind. *Royal Society Open Science, 9*(10), Article 211278. https://doi.org/10.1098/rsos.211278

Berke, M., Horschler*, D., Royka, A., Santos, L., & Jara-Ettinger, J. (2025). What primates know about other minds and when they use it: A computational approach to comparative theory of mind. *biorxiv*.

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric Bias in mental-state attribution. *Psychological Science, 14*(3), 283–286. https://doi.org/10.1111/1467-9280.03436

Bond, A. B., Kamil, A. C., & Balda, R. P. (2003). Social complexity and transitive inference in corvids. *Animal Behaviour, 65*(3), 479–487. https://doi.org/10.1006/anbe.2003.2101

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature, 232*(5311), 456–458. https://doi.org/10.1038/232456a0

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), Article e0173793. https://doi.org/10.1371/journal.pone.0173793

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

Cantlon, J. F., & Brannon, E. M. (2007). How much does number matter to a monkey (*Macaca mulatta*)? *Journal of Experimental Psychology: Animal Behavior Processes, 33*(1), 32–41. https://doi.org/10.1037/0097-7403.33.1.32

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28*(2), 141–172. https://doi.org/10.1111/mila.12014

Cordes, S., & Brannon, E. M. (2009). The relative salience of discrete and continuous quantity in young infants. *Developmental Science, 12*(3), 453–463. https://doi.org/10.1111/j.1467-7687.2008.00781.x

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology, 22*(2), 142–146. https://doi.org/10.1016/j.cub.2011.11.053

Crockford, C., Wittig, R. M., & Zuberbühler, K. (2017). Vocalizing in chimpanzees is influenced by social-cognitive processes. *Science Advances, 3*(11), Article e1701742. https://doi.org/10.1126/sciadv.1701742

Davis, H. (1992). Transitive inference in rats (*Rattus norvegicus*). *Journal of Comparative Psychology, 106*(4), 342–349. https://doi.org/10.1037/0735-7036.106.4.342

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. https://doi.org/10.1016/j.cogdev.2018.01.001

Drayton, L. A., & Santos, L. R. (2017). Do rhesus macaques, *Macaca mulatta*, understand what others know when gaze following? *Animal Behaviour, 134*, 193–199. https://doi.org/10.1016/j.anbehav.2017.10.016

Drayton, L. A., & Santos, L. R. (2018). What do monkeys know about others' knowledge? *Cognition, 170*, 201–208. https://doi.org/10.1016/j.cognition.2017.10.004

Engelmann, J. M., Herrmann, E., Proft, M., Keupp, S., Dunham, Y., & Rakoczy, H. (2022). Chimpanzees consider freedom of choice in their evaluation of social action. *Biology Letters, 18*(2), Article 20210502. https://doi.org/10.1098/rsbl.2021.0502

Fabricius, W. V., Gonzales, C. R., Pesch, A., Weimer, A. A., Pugliese, J., Carroll, K., … Spinrad, T. L. (2021). Perceptual access reasoning (PAR) in developing a representational theory of mind. *Monographs of the Society for Research in Child Development, 86*(3), 7–154. https://doi.org/10.1111/mono.12432

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*(5), 568–584. https://doi.org/10.1111/1467-7687.00313

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(3), 295–313. https://doi.org/10.1016/j.cognition.2004.09.010

von Fersen, L., Wynne, C. D., Delius, J. D., & Staddon, J. E. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 17*(3), 334–341. https://doi.org/10.1037/0097-7403.17.3.334

Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology, 15*(5), 447–452. https://doi.org/10.1016/j.cub.2004.12.076

Garnham, W. A., & Perner, J. (2001). Actions really do speak louder than words—But only implicitly: Young children's understanding of false belief in action. *British Journal of Developmental Psychology, 19*(3), 413–432. https://doi.org/10.1348/026151001166182

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*(1), 1–14. https://doi.org/10.1017/S0140525X00028636

Grosse Wiesmann, C., Friederici, A. D., Disla, D., Steinbeis, N., & Singer, T. (2018). Longitudinal evidence for 4-year-olds' but not 2- and 3-year-olds' false belief-related action anticipation. *Cognitive Development, 46*, 58–68. https://doi.org/10.1016/j.cogdev.2017.08.007

Hamlin, K. J., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science, 16*(2), 209–226. https://doi.org/10.1111/desc.12017

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour, 59*(4), 771–785. https://doi.org/10.1006/anbe.1999.1377

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*(1), 139–151. https://doi.org/10.1006/anbe.2000.1518

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., … Hasegawa, I. (2020). Macaques exhibit implicit gaze Bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Reports, 30*(13), 4433–4444. e5 https://doi.org/10.1016/j.celrep.2020.03.013.

Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*(3), 567–582. https://doi.org/10.2307/1130337

Holland, C., & Phillips, J. (2020). A theoretically drivem meta-analysis of implicit thoery of mind studies: The role of factivity. *Proceedings of the Annual Meeting of the Cognitive Science Society, 42*.

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

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

Horschler, D. J., Santos, L. R., & MacLean, E. L. (2021). How do non-human primates represent others' awareness of where objects are hidden? *Cognition, 212*, Article 104658. https://doi.org/10.1016/j.cognition.2021.104658

Howard, S. R., & Barron, A. B. (2024). Understanding the limits to animal cognition. *Current Biology, 34*(7), R294–R300. https://doi.org/10.1016/j.cub.2024.02.043

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

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

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

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015b). The goggles experiment: Can chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour, 105*, 211–221. https://doi.org/10.1016/j.anbehav.2015.04.028

Krachun, C., Carpenter, C. M., Call, J., & Tomasello, M. (2010). A new change-of-contents false belief test: Children and chimpanzees compared. *International Journal of Comparative Psychology, 23*(2). https://doi.org/10.46867/ijcp.2010.23.02.03

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

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

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

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cognitive Development, 46*, 97–111. https://doi.org/10.1016/j.cogdev.2017.09.001

Lewis, K. P., Jaffe, S., & Brannon, E. M. (2005). Analog number representations in mongoose lemurs (*Eulemur mongoz*): Evidence from a search task. *Animal Cognition, 8*(4), 247–252. https://doi.org/10.1007/s10071-004-0251-x

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

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science, 12*(1), 142–149. https://doi.org/10.1111/j.1467-7687.2008.00741.x

Lurz, R., Samuel, S., Mshvelidze, J., Cherry, A., Dent, J., Garcia, J., … Taglialatela, J. (2025). Are bonobos anxious when others are ignorant of a threat? Using an affect measure to assess theory-of-mind abilities in bonobos. *Psihologijske Teme, 34*(1), 139–154. https://doi.org/10.31820/pt.34.1.7

Lurz, R. W., Krachun, C., Mareno, M. C., & Hopkins, W. D. (2022). Do chimpanzees predict others' behavior by simulating their beliefs? *Animal Behavior and Cognition, 9*(2), 153–175. https://doi.org/10.26451/abc.09.02.01.2022

MaBouDi, H., Barron, A. B., Li, S., Honkanen, M., Loukola, O. J., Peng, F., … Solvi, C. (2021). Non-numerical strategies used by bees to solve numerical cognition tasks. *Proceedings of the Royal Society B: Biological Sciences, 288*(1945), Article 20202711. https://doi.org/10.1098/rspb.2020.2711

MacLean, E. L., & Hare, B. (2012). Bonobos and chimpanzees infer the target of another's attention. *Animal Behaviour, 83*(2), 345–353. https://doi.org/10.1016/j.anbehav.2011.10.026

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

Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition, 130*(3), 300–308. https://doi.org/10.1016/j.cognition.2013.11.016

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

McGonigle, B., & Chalmers, M. (1984). The selective impact of question form and input mode on the symbolic distance effect in children. *Journal of Experimental Child Psychology, 37*(3), 525–554. https://doi.org/10.1016/0022-0965(84)90075-4

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology, 120*, 154–162. https://doi.org/10.1037/0735-7036.120.2.154

Mossler, D. G., Marvin, R. S., & Greenberg, M. T. (1976). Conceptual perspective taking in 2- to 6-year-old children. *Developmental Psychology, 12*(1), 85–86. https://doi.org/10.1037/0012-1649.12.1.85

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

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., & Knobe, J. (2021). Knowledge before belief. *Behavioral and Brain Sciences, 44*, Article e140. https://doi.org/10.1017/S0140525X20000618

Phillips, J., & Norby, A. (2021). Factive theory of mind. *Mind & Language, 36*(1), 3–26. https://doi.org/10.1111/mila.12267

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., … Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet – A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development, 48*, 302–315. https://doi.org/10.1016/j.cogdev.2018.09.005

Poulin-Dubois, D., & Yott, J. (2018). Probing the depth of infants' theory of mind: Disunity in performance across paradigms. *Developmental Science, 21*(4), Article e12600. https://doi.org/10.1111/desc.12600

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. https://doi.org/10.1016/j.cogdev.2017.10.004

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Rosati, A. G., & Santos, L. R. (2016). Spontaneous metacognition in Rhesus monkeys. *Psychological Science, 27*(9), 1181–1191. https://doi.org/10.1177/0956797616653737

Rugani, R., Regolin, L., & Vallortigara, G. (2010). Imprinted numbers: Newborn chicks' sensitivity to number vs. continuous extent of objects they have been reared with. *Developmental Science, 13*(5), 790–797. https://doi.org/10.1111/j.1467-7687.2009.00936.x

Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, *Macaca mulatta*, know what others can and cannot hear. *Animal Behaviour, 71*(5), 1175–1181. https://doi.org/10.1016/j.anbehav.2005.10.007

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80*(4), 1172–1196. https://doi.org/10.1111/j.1467-8624.2009.01324.x

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), 1239–1247. https://doi.org/10.1037/a0023899

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

Sullivan, K., & Winner, E. (1991). When 3-year-olds understand ignorance, false belief and representational change. *British Journal of Developmental Psychology, 9*(1), 159–171. https://doi.org/10.1111/j.2044-835X.1991.tb00868.x

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology, 30*(1), 30–44. https://doi.org/10.1111/j.2044-835X.2011.02046.x

Taylor, A. H., Bastos, A. P. M., Brown, R. L., & Allen, C. (2022). The signature-testing approach to mapping biological and artificial intelligences. *Trends in Cognitive Sciences, 26*(9), 738–750. https://doi.org/10.1016/j.tics.2022.06.002

Townrow, L. A., & Krupenye, C. (2025). Bonobos point more for ignorant than knowledgeable social partners. *Proceedings of the National Academy of Sciences, 122*(6), Article e2412450122. https://doi.org/10.1073/pnas.2412450122

Treichler, F. R., & Van Tilburg, D. (1996). Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List linking. *Journal of Experimental Psychology: Animal Behavior Processes, 22*(1), 105–117. https://doi.org/10.1037/0097-7403.22.1.105

vanMarle, K., Aw, J., McCrink, K., & Santos, L. R. (2006). How capuchin monkeys (*Cebus apella*) quantify objects and substances. *Journal of Comparative Psychology, 120*(4), 416–426. https://doi.org/10.1037/0735-7036.120.4.416

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science, 17*(6), 872–879. https://doi.org/10.1111/desc.12170

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9