

Opinion

Do Non-Human Primates Really Represent Others' Beliefs?

Daniel J. Horschler,^{1,2,*} Evan L. MacLean,^{1,2,3,4} and Laurie R. Santos⁵

Over two decades of research have produced compelling evidence that non-human primates understand some psychological states in other individuals but are unable to represent others' beliefs. Recently, three studies employing anticipatory looking (AL) paradigms reported that non-human primates do show hints of implicitly understanding the beliefs of others. However, measures of AL have been increasingly scrutinized in the human literature owing to extensive replication problems. We argue that new reports of belief representation in non-human primates using AL should be interpreted cautiously because of methodological and theoretical challenges paralleling trends in the human literature. We explore how future work can address these challenges, and conclude by identifying new evolutionary questions raised by the prospect that non-human primates implicitly represent others' beliefs without an explicit belief representation system that guides fitness-relevant behavior.

Is Belief Representation Human-Unique?

Theory of mind (ToM; see [Glossary](#)) is a fundamental aspect of human cognition [1]. One component of ToM, **false belief** (FB) representation, has long been considered to be the hallmark of human mental state reasoning [2–4]. FB **representation** allows us to understand that others can believe things that are different from our own understanding of reality. FB representation follows a clear developmental trajectory in humans, with explicit understanding of FBs (i.e., **explicit belief representation**) emerging ~4 years of age [5–9]. However, several studies have also suggested that an implicit understanding of FBs (i.e., **implicit belief representation**) may develop within the first 2 years of life [10–13].

In the same way as developmental psychologists have tested how FB representations emerge in development, comparative researchers have investigated whether our closest non-human relatives share such representational capacities ([14–17] for reviews). These studies have largely shown that non-human primates (hereafter primates) fail to represent the FBs of others, both when tested in explicit choice tasks [18–22] and implicitly via **expectancy violation** [23,24]. However, three recent papers [25–27] tested primates using a different implicit measure – **anticipatory looking** (AL) – and reported the first comparative evidence of FB representation (*cf* [28,29]). The results of these studies are exciting, and the development of AL methods for primates opens the door to a host of new possibilities for research on the evolution of ToM. Unfortunately though, human AL studies of FB representation have suffered from major replicability problems, leading to debate about the validity of these methods. Thus, although the application of AL to research on non-human animals is an important innovation with tremendous potential for future studies, caution is warranted given that (i) the validity of these measures has been challenged in human research, and (ii) AL studies of belief representation with primates have produced results that contradict a long history of findings from more conventional paradigms. We address here the theoretical

Highlights

A large body of research supports the idea that non-human primates are unable to represent others' beliefs.

Three recent studies challenge this view and report evidence of implicit belief representation in non-human primates using AL measures that track eye gaze as an index of expectation about others' actions.

These findings should be interpreted cautiously owing in part to replication issues with AL studies of implicit belief representation in humans. We identify parallel methodological and theoretical challenges in comparative work.

In human development, implicit understanding of others' beliefs precedes explicit representation, and the latter guides intentional behavior. That non-human primates might represent others' beliefs implicitly, but not explicitly, raises challenging evolutionary questions: a system for implicit belief representation that does not guide fitness-relevant behavior would remain invisible to direct evolutionary selective pressure.

¹School of Anthropology, University of Arizona, Tucson, AZ 85719, USA

²Cognitive Science Program, University of Arizona, Tucson, AZ 85719, USA

³Department of Psychology, University of Arizona, Tucson, AZ 85719, USA

⁴College of Veterinary Medicine, University of Arizona, Tucson, AZ 85719, USA

⁵Department of Psychology, Yale University, New Haven, CT 06520, USA

*Correspondence: horschler@email.arizona.edu (D.J. Horschler).

©Twitter: @DanielHorschler



and methodological questions raised by these new results, synthesizing across the developmental and comparative literatures.

Tests of Belief Representation in Primates

Decades of work have supported the conclusion that primates are incapable of belief representation (Table 1, Key Table). In tests of FB representation, researchers typically construct scenarios in which subjects should perform differently depending on whether they are reacting to what they themselves know to be true, or to what another agent incorrectly believes. In **change-of-location tasks**, a subject and agent watch as an object is hidden among multiple locations, and is then removed from its initial location and hidden elsewhere while the agent's view is occluded. In these scenarios, the agent thus has a FB about the object's location while the subject maintains **knowledge** based on their **true belief** (TB).

Comparative change-of-location tasks have typically used subjects' behavioral responses (e.g., reaching) as the dependent measure. Non-human apes (hereafter apes) have consistently failed to differentiate how agents with FBs would act relative to control conditions in both cooperative [18] and competitive [19,20,22] contexts, despite generally correctly predicting the actions of knowledgeable agents. Apes showed this same pattern in a change-of-contents task, in which the identity of a hidden object was changed [21]. In only two studies (to our knowledge) have authors argued for hints of explicit FB representation in apes [28,29], but in both cases apes' performance in knowledge and/or ignorance control conditions suggested that belief representation was unlikely to explain their behavior (section on Differentiating Knowledge, Ignorance, and Belief).

By contrast, implicit tasks use looking patterns rather than explicit behavioral responses to assess understanding of others' mental states. In one implicit measure – expectancy violation – subjects are predicted to look longer at events that are incongruent with their expectations. Two studies [23,24] using designs adapted from the human literature [10,11] found that rhesus macaques successfully recognize when others have knowledge, but are not sensitive to others' FBs. One experiment found that apes oriented their face toward the unchosen location more often in a FB than knowledge condition, but this result failed to replicate in a second experiment that directly assessed eye direction [20].

In contrast to numerous failures to demonstrate belief representation in primates, three studies [25–27] recently reported that primates pass change-of-location FB tests using implicit AL measures, where looking patterns are believed to indicate a subject's expectation about the location an agent will next act on (*cf* [30] for preliminary negative AL results with rhesus macaques). These innovative studies circumvent many challenges inherent to previous explicit and implicit paradigms, and should be applauded for pioneering the use of IR eye tracking in comparative ToM tasks. However, given the nascent stage of this work and its potential for reshaping the future of comparative ToM research, it is crucial to understand why primates consistently fail tests of belief representation measured via behavioral responses and expectancy violation, but pass some tests via AL [25–27].

Replicability and Robustness of AL in the Human Literature

The authors of the comparative AL papers [25–27] argued that AL provides a reliable and robust measure of subjects' expectations about others' beliefs, but recent developmental studies call this into question. AL has indeed been used to assess FB representation in many studies of human infants, children, and adults [31–61], but taken together this work reveals that AL measures of FB representation are neither stable across the lifespan [47] nor highly replicable

Glossary

Anticipatory looking (AL): an implicit measure used in tests of belief representation. The location a subject looks to before an event is believed to reflect their implicit expectation of what will occur.

Change-of-location task: an experimental paradigm used to induce false beliefs (FBs). In these tasks, a subject and agent watch as an object is hidden among multiple locations, but is then removed from its initial hiding location and hidden elsewhere while the agent's view is occluded.

Differential looking score (DLS): a statistical measure of a subject's AL pattern. DLS is calculated by subtracting total looking time to the incorrect location from total looking time to the correct location, and then dividing by the sum of these times.

Expectancy violation: an implicit measure used in tests of belief representation. The time a subject spends looking at the location of a recent event is believed to reflect their prior predictions about that event, such that longer looking times indicate incongruence between what the subject predicted and what occurred.

Explicit belief representation: a consciously accessible representation of an agent's belief that can influence intentional behavior. Tasks measuring explicit belief representation often require subjects to provide a verbal response (in humans) or a behavioral response such as reaching (in primates).

False beliefs (FBs): beliefs that are incongruent with reality, including instances where an agent's belief is incongruent with a subject's representation of what is true.

Gettier problem: a philosophical argument that knowledge is not equivalent to justified true belief (TB), because the reasons behind an agent's justified TB may be false or incomplete.

Implicit belief representation: an automatic or unconscious understanding of an agent's belief. In both humans and primates, implicit belief representation tasks use a subject's looking patterns to infer their prediction about an agent's future action (AL) or their reaction when their prediction is or is not fulfilled (expectancy violation).

Justified true beliefs: beliefs about true information that are formed for appropriate reasons from a subject's

Key Table

Table 1. Tests of Belief Representation in Non-human Primates

Report	Measure	Main results
Call and Tomasello [18]	Explicit	In a cooperative paradigm, apes incorrectly reached toward a hiding location indicated by a human communicator who falsely believed that food was hidden there, suggesting that apes were not sensitive to the veracity of the communicator's belief. Apes passed all control conditions, indicating that failure was unlikely to be due to task demands.
Hare <i>et al.</i> [22]	Explicit	Apes made no distinction between conspecific competitors who were either ignorant or had a FB about the location of hidden food despite understanding how knowledgeable competitors would act.
O'Connell and Dunbar [29]	Explicit	Apes showed some sensitivity to a human experimenter's FB about the location of a marker that indicated where the experimenter should hide a food reward, but crucially failed a corresponding knowledge condition (section on Differentiating Knowledge, Ignorance, and Belief).
Kaminski <i>et al.</i> [19]	Explicit	Apes showed no evidence of understanding conspecific competitors' FBs about the location of hidden food (relative to control conditions) despite successfully acting on competitors' knowledge. When a competitor's view was occluded, apes made no distinction between instances where the hidden food was simply lifted out of its hiding location and placed back into the same location (justified TB) versus instances where the food was moved to a new location (FB).
Krachun <i>et al.</i> [20]	Explicit	In a competitive paradigm, apes incorrectly chose a hiding location effortfully reached toward by a human competitor who falsely believed food was hidden there, despite passing an analogous condition where the competitor was knowledgeable.
	Implicit	Apes oriented their face toward the unchosen container more often in a FB than knowledge condition, potentially indicating some sensitivity to the competitor's incongruent action, but this result failed to replicate in a second experiment that directly assessed eye direction instead of using face direction as a proxy.
Krachun <i>et al.</i> [21]	Explicit	In a change-of-contents versus change-of-location task, where the identity instead of the location of a hidden object was manipulated, apes showed no evidence of understanding when a human experimenter had a FB about the identity of the object, despite understanding how a knowledgeable experimenter would act.
Marticoirena <i>et al.</i> [23]	Implicit	Rhesus macaques looked significantly longer when a human experimenter with knowledge about hidden food reached for the incorrect rather than correct location, but showed no difference in looking time when the experimenter had a FB, indicating no expectation about the behavior of an agent with a FB.
Martin and Santos [24]	Implicit	In an automatic belief representation task, rhesus macaques looked significantly longer at events that violated their own beliefs, but not those of a human experimenter, about where food was hidden.
Krupenye <i>et al.</i> [25]	Implicit	Apes showed some evidence of AL that was consistent with accurate predictions about where a human experimenter with a FB would search for a hidden object when data from two different designs across two different experiments were combined (section on Methodological Challenges in Comparative Approaches to AL).
Buttelmann <i>et al.</i> [28]	Explicit	Apes correctly helped a human experimenter to open a box containing a desired object when the experimenter attempted to open a different box that he falsely believed the object to be in, but crucially failed a corresponding knowledge condition (section on Differentiating Knowledge, Ignorance, and Belief).
Horschler <i>et al.</i> [86]	Implicit	After a human experimenter witnessed where food was hidden, and the food was then moved out of and back into the same hiding location while the experimenter's view was occluded, rhesus macaques attributed no justified TB about the food's location to the experimenter – they showed no difference in looking time at the experimenter's correct versus incorrect reaches. Rhesus macaques looked significantly longer at incorrect reaches in a knowledge condition where an irrelevant aspect of the scene was instead manipulated while the experimenter's view was occluded.
Kano <i>et al.</i> [26]	Implicit	Apes showed some evidence of AL that was consistent with accurate predictions

point of view, even if these reasons turn out to be false or incomplete.

Knowledge: at least in the context of primate FB tests, an understanding of true information based on direct perceptual access to all information leading to the current state of events.

Reality bias: an expectation that an agent will act based on what is true (also known as a 'curse of knowledge' or 'true belief default').

Representations: cognitive abstractions of information, such as physical or intangible objects or concepts. In belief representation, a subject must represent others' representations of the world.

Theory of mind (ToM): the ability to represent the mental states of others, such as knowledge, ignorance, beliefs, and desires.

True beliefs (TBs): beliefs that are congruent with reality.

Table 1. (continued)

Report	Measure	Main results
		about where a human experimenter with a FB would search for a hidden object in a context similar to Krupenye <i>et al.</i> [25], but using a different primary dependent measure (section on Methodological Challenges in Comparative Approaches to AL).
Hayashi <i>et al.</i> [27]	Implicit	Japanese macaques showed some evidence of AL that was consistent with accurate predictions about where a human experimenter with a FB would search for a hidden object when data were combined from two different designs across two scenarios similar to Krupenye <i>et al.</i> [25] and a third scenario involving a disinterested puppet. However, different dependent measures supported different conclusions in some conditions (section on Methodological Challenges in Comparative Approaches to AL).

(Table 2, and Table S1 in the supplemental information online), and non-replications appear in all age groups [47–53,62–64]. Although some successful replications have been reported (Table S1), other researchers have noted additional unpublished non-replications, suggesting the possibility of publication bias [64], which is also supported by a recent meta-analysis of this literature [65]. One published example replicated only one of four studies in adults despite using the original stimuli and procedures, and the sole successful replication also failed once an important confound was eliminated [51]. In addition, there was no evidence of convergent validity across any of the four AL tasks, raising serious concerns about whether AL can be used to measure belief representation [51]. Similarly, results from AL measures have proved to be incongruent with adult subjects' verbal explanations of their action anticipations [45]. Other non-replications have emerged when larger sample sizes [47,49–52] and more socially engaging stimuli [53] were used. Thus, even in adult humans, where FB representation can be confirmed using explicit measures, AL has presented a host of challenges that remain to be resolved. Although replication issues have also emerged with expectancy violation measures of FB representation [48,62,63,66,67], procedural differences may plausibly account for many of these reports [62]. Non-replications of AL tasks are both more numerous and more troubling given that many used the original stimuli (e.g., videos) in their replication attempts [62], and a recent meta-analysis found a significant mean effect size for expectancy violation, but not for AL tasks [65].

In addition to failed strict replications of AL tasks [48–51,53,62–64], researchers have also observed conceptual replication issues, perhaps arising from methodological inconsistencies across studies ([68] for similar critiques of expectancy violation studies). For example, some studies reported first look as the primary dependent measure [31,34,36,39,40,45–48,50,52–54,58,59], whereas others used **differential looking score** (DLS) [33,37,38,41,42,49–51,55–57], total looking time [32,43,61], or proportion of total looking towards the correct location [35,44,51,60], and many studies employed more than one of these measures [34,36,38–42,45–54,56–59]. In addition, there is considerable variation in exclusion criteria based on familiarization trial performance [34–36,38–42,44,45,47–54,56,59,60] (Table S1). Some studies justify these criteria by arguing that correct action anticipation on familiarization trials is informative because these trials are similar to simple knowledge conditions [54], whereas others argue that familiarization trials are necessary to orient subjects to novel stimuli before they are able to generate action predictions [25]. In any case, passing or failing familiarization trials appears to have no relationship to test trial performance at an individual level [49,50,63]. Variation is also extensive in areas of interest that denote where looks are meaningful, and response time-windows for looking analyses have included disparate values ranging from as short as 0.6 s to as long as 6 s (Table S1). Although some studies justify the length of this response window by matching it to the pause duration before the agent reaches in familiarization trials [35,54,56,57,59], often no theoretical rationale is provided.

In sum, these non-replications of AL tasks provide no clear pattern of results regarding which procedural variants are most valid [62]. Therefore, although AL measures may hold promise in elucidating mental state representation in nonverbal subjects, it is currently difficult to know whether and which AL outcomes reliably capture belief representations [53], posing challenges for translation to comparative studies.

Methodological Challenges in Comparative Approaches to AL

Comparative studies using AL also face challenges with respect to both study design and the choice of dependent measures. AL studies in humans have used a variety of designs to induce FBs in agents [48–51,62–64]. All three comparative AL papers [25–27] were modeled on a design by Southgate and colleagues [54], which unfortunately has a relatively high non-replication rate compared to other designs (Table 2; Table S1 lists successful replications). It is also noteworthy that the stimuli used in comparative studies sometimes differ, and this could influence replicability in these populations. The authors of the ape papers argued that the FB2 condition of the Southgate design imposes 'a more stringent test of action anticipation' specifically 'because recent attempts to replicate Southgate *et al.* with human populations found greater difficulty replicating the FB2 design' [26] (see Box 1 for methods). However, other scholars claim that this design's replication problems reflect poor indexing of action anticipation and belief representation [63]. In a commentary on these issues, Southgate herself concluded: 'It seems fairly clear ... that the original paradigm does not consistently elicit or reveal spontaneous action prediction or epistemic-state tracking across multiple age ranges and populations' [62], raising concerns about its suitability for comparative studies.

Table 2. Partial and Non-replications of AL Belief Representation Tests in Humans^a

	Report	Mean sample age (years)	Test trial response window(s)	Dependent measure(s)	Familiarization exclusion criteria
Original paradigm	Southgate <i>et al.</i> [54]	~2	1.75 s	First look; total looking time	AL on the second of two familiarization trials
Partial and non-replications: Southgate <i>et al.</i>	[40,42,47,48], [49] ^{1,2} , [50] ^{1a} , [50] ^{1b} , [50] ^{2a} , [50] ^{2b} , [51] ^{1d} , [52] ¹ , [52] ² , [53] ¹ , [53] ² , [57]	~2–10.5 Adult	1.75–6 s	First look; total looking time; proportion of total looking to correct location; DLS	Same: [40,48], [49] ¹ , [50] ^{1a} , [50] ^{1b} , [52] ¹ , [52] ² , [53] ¹ , [53] ² Different: [40,42,47], [49] ^{1,2} , [50] ^{1a} , [50] ^{2a} , [50] ^{2b} , [51] ^{1d} , [57]
Original paradigm	Schneider <i>et al.</i> [61]	Adult	5 s	Total looking time	None
Partial and non-replications: Schneider <i>et al.</i>	[51] ^{1a} , [51] ²	Adult	5 s	First look; proportion of total looking to correct location	Same: [51] ^{1a} , [51] ² Different: none
Original paradigm	Surian and Geraci [34]	~1.5 Adult	3.5 s	First look; total looking time	AL on either of two familiarization trials
Partial and non-replications: Surian and Geraci	[50] ^{2b} , [51] ^{1b} , [53] ¹ , [53] ²	~4.5–5.5 Adult	3.5–6 s	First look; proportion of total looking to correct location	Same: [50] ^{2b} , [51] ^{1b} Different: [53] ¹ , [53] ²
Original paradigm	Low and Watts [36]	~3–4 Adult	1.75 s	First look; DLS	AL on the second of two familiarization trials
Partial and non-replications: Low and Watts	[51] ^{3a} , [51] ^{3b}	Adult	1.75 s	DLS	Same: [51] ^{3a} , [51] ^{3b} Different: none

^aSuperscripts indicate specific experiments within a paper. Detailed methods and main results from each experiment cited above are summarized in Table S1 together with any successful replication attempts of these paradigms. All experiments used a change-of-location task. Some of the experiments cited above are conceptual rather than strict replication attempts that include intentional methodological departures from the original studies.

Box 1. Methods in Comparative AL Studies

In three recent comparative FB tests [25–27], subjects watched videos during which a human acquired a FB while subjects' AL was recorded using an IR eye-tracker. Krupenye and colleagues [25] ran two experiments with identical designs, but superficially dissimilar scenarios. In this study, a human pursued a target object that could be hidden in one of two places. In two familiarization trials intended to demonstrate the human's goal, the human watched the object being hidden in each location and immediately pursued it. In the test trials, the human first witnessed the object being hidden in one location, and then did not see as it was either removed from the scene (FB1 condition) or moved to the other location and then removed from the scene (FB2 condition). Therefore, in both conditions of each experiment, the human had a FB about the location of an object because it never remained in the scene. Subsequently, apes' AL was recorded as the human centrally approached both locations. Experiment 1 involved the human searching for an ape-like character (KK) who hid in one of two haystacks, whereas experiment 2 involved the human searching for a stone that KK stole and hid in one of two boxes. Hayashi and colleagues [27] ran analogous FB1 and FB2 conditions with Japanese macaques across three superficially dissimilar scenarios: two humans aggressively competing over an apple, one human chasing another, and a disinterested puppet manipulating a toy with a human onlooker.

Kano and colleagues [26] used the same design and scenario as the FB2 condition in experiment 2 of [25], but with a change modeled after the 'goggles test' [78,79,114–117]. Before the experiment, apes were introduced to a barrier which appeared to be opaque from a distance but was in reality either opaque or transparent when viewed up close. Between subjects, apes learned that the barrier was either opaque or transparent, and then all subjects watched the same series of events: the human witnessed the target object being hidden in one location, and then moved behind the barrier as the object was moved to the other location before finally being removed from the scene. Thus, apes who previously learned that the barrier was transparent should have expected the human to have knowledge that the object was no longer present, whereas apes who learned that the barrier was opaque should have expected the human to have a FB about the object's location.

Another methodological issue with the comparative papers involves the use of different primary dependent measures. The first ape study [25] tested subjects' first looks in a 4.5 s response time-window, but the second ape study [26] primarily used DLS in a 6 s response window despite using the same design as a condition in [25]. Although secondary analyses of total looking time showed consistent results with first looks in [25], the performance of apes in [26] suggested discrimination between TB and FB conditions when analyzing DLS, but not first looks. One study of Japanese macaques [27] reported first looks and DLS as primary dependent measures in response time-windows ranging from 8 to 11.46 s. Although DLS and first looks showed consistent results in one condition (FB2), first looks but not DLS suggested belief representation in another (FB1). In addition, after a manipulation that deactivates the medial prefrontal cortex, first looks but not DLS suggested significantly different effects across conditions. The different dependent measures used in these studies therefore often support different conclusions, a pattern also observed in the human literature [34,47,48,50–53,57].

Apes also performed differently across theoretically identical experiments, even when using the same dependent measure. In [25], apes looked first at the target significantly more than at the distractor (suggestive of FB representation) in experiment 2 but not in experiment 1 (although the results trended in the same direction). In addition, in the opaque-only condition of [26] – which was virtually identical to experiment 2 in [25] – apes did not look first at the target significantly more than the distractor, suggesting limited replicability using this measure. This variability in results across highly similar experiments is reminiscent of the many partial and non-replications observed in human populations [48–51,53,62–64]. A reanalysis of experiment 2 from [25] using DLS instead of first looks did replicate DLS results from the opaque-only condition of [26], but importantly this analysis was presented in [26] rather than in the original paper [25]. Thus, although retrospective analyses can identify dependent measures that yield similar results, this is perhaps unsurprising given the range of possible measures to be employed, and variation in which specific measures are emphasized. Lastly, many apes entirely failed to make anticipatory looks [25,26], and both apes and macaques performed inconsistently at the individual level [25,27], raising additional questions about the robustness of AL measures (Box 2).

Box 2. Individual Differences in Action Anticipation

In AL paradigms, low rates of looking may indicate a lack of clear predictions about what an agent will do next in cases where subjects attend to all belief induction events. In [25], eight of 29 apes made no anticipatory looks in each of the two experiments, and in [26], 17 of 47 apes made no anticipatory looks. This is in stark contrast to the original study of 2-year-old children using this design [54], in which only two of 36 children made no anticipatory looks during the test trial despite using a shorter 1.75 s response time-window instead of 4.5 s in [25] or 6 s in [26] ([48] for similar results in 2-year-olds). Importantly, because all apes in [25,26] were tested in an environment with significantly higher potential for distraction owing to nearby conspecifics, it is unclear to what extent this pattern reflects inattention versus a lack of clear action predictions. Although the authors did confirm sufficient attention to belief induction events in [25,26], future work should quantify distraction and inattention during test trials to better elucidate the reasons behind apes' lower AL rates. Hayashi and colleagues [27] reported a slightly lower no-look rate in macaques, possibly because of the use of restraints (i.e., primate chairs) or longer response time-windows (8–11.46 s) than in any previous AL study.

Similarly, if AL reliably indexes belief representation, we should expect convergent validity such that patterns of individual differences are consistent across tests. In humans, stable individual differences in verbal ToM task performance are evident across development [8,9], and convergent validity of superficially dissimilar explicit ToM tasks is well documented [51,118]. However, adult humans show no significant correlations between performance on superficially dissimilar AL FB tests at the individual level [51]. Primates appear to show a similar trend. In [25], only eight of 29 apes passed both versions of the FB test, whereas nine passed one but failed the other. In [27], macaques performed consistently on left and right counterbalanced versions of the same test in only nine instances (with two consistently failing), whereas in another nine instances a subject passed one version but failed the other. Although more work will be necessary to robustly assess individual differences in primate ToM, these findings challenge the notion that AL provides a reliable index of belief state tracking in either primates or humans.

Key Features of Comparative AL Designs

The comparative AL papers [25–27] argued that primates' previous failures in belief representation tests [18–24] but success in AL tasks may have resulted partially from the authors' use of socially dramatic scenes (Box 1) which may better capture primates' attention than the scenarios presented in previous studies. However, one AL study that replaced the socially dramatic scene with differently colored shapes [69] found that apes closely tracked all belief induction events just as in [25], suggesting that social drama is not necessary to adequately capture apes' attention. Similarly, Hayashi and colleagues [27] included one scenario enacted by a disinterested puppet, and found a greater proportion of correct first looks in this scenario compared to two socially dramatic scenarios. These results suggest that socially dramatic stimuli may not be key to eliciting belief-tracking in primates.

Baillargeon and colleagues [12,62] argued that forming expectations about others' belief-based actions may be taxing and relatively time-intensive for nonverbal subjects, whereas AL assumes that action predictions can be made swiftly and accurately. Theoretically, online mental state tracking in AL should then be more cognitively demanding than post-event mental state tracking, as in the expectancy violation FB tests that macaques have previously failed [23,24]. Unpublished results from Joiner [30] replicate the expectancy violation findings of [23] but show no AL in the knowledge condition and very low rates of AL, with no significant bias toward either location in the FB condition. Therefore, these preliminary results suggest that, in a scenario lacking social drama, macaques show sensitivity to an agent's knowledge but not FBs when measured via expectancy violation but not AL.

Although it thus appears unlikely that social drama is fully responsible for primates' success in [25–27], it is possible that apes failed previous explicit tests because of task demands [70] (e.g., inhibitory control demands in paradigms involving food). However, apes have demonstrated strong self-control across a range of contexts [71–75], exceeding that of young children [76] and sometimes even of adult humans [77], making it unlikely that inhibitory control demands fully account for apes' failures on explicit FB tests.

Theoretical Interpretations of Belief Representation Tests

Separate from the methodological issues discussed above, there is also considerable debate surrounding the interpretation of belief representation tests. This debate centers on whether positive results from AL and other belief representation tests in both humans and primates reflect mental state representation as such, or instead reflect other cognitive mechanisms that allow belief-congruent inferences about goal-directed action without belief state representation. Indeed, many scholars have proposed alternative explanations including behavior-reading [78], submentalizing [79], minimal ToM [80], experiential record-keeping [81], and awareness relations [82]. Therefore, positive results on any given belief representation test may not necessarily indicate belief state representation as such [83,84].

Many primate studies have focused on directly testing FB representation (see above). However, TB tests can also provide unique insight into whether primates truly represent others' beliefs. More specifically, TB tests modeling the **Gettier problem** [85] have shown that primates' understanding of others' TBs is not always robust. The Gettier problem is a special case in philosophy in which an agent ends up having a **justified true belief** about a given situation (e.g., they believe an object to be where it actually is) despite lacking information relevant to actually knowing about what they believe (e.g., they were unaware that the object was moved several times beforehand). Kaminski *et al.* [19] and Horschler *et al.* [86] found that apes and macaques no longer expect an agent to have a previously acquired TB about an object's location after it is arbitrarily moved out of, and then back into, the same hiding location while the agent's view is occluded. In these studies, because the agent does not see the object's final spatial manipulation, they cannot be said to have knowledge of the object's location (only a TB). If primates represent others' beliefs as such, this arbitrary manipulation should not disrupt their predictions about the agent's action given that it does not impact the veracity of the agent's belief. Coupled with failure on explicit FB tests, failure on these TB tests therefore suggests that primates do not represent others' belief states as such [82].

Differentiating Knowledge, Ignorance, and Belief

Knowledge and ignorance control conditions in experiments testing FB representation are crucial because they should yield differentiable results from each other and from FB conditions. Even without representing belief or ignorance states as such [82], it is possible that subjects may pass some FB conditions by predicting that agents who lack complete information about an object's location tend to search incorrectly [34,87]. Although the Southgate design attempts to control for this possibility by rendering both locations incorrect, it introduces additional challenges by precluding knowledge conditions with clear associated predictions (see below). In change-of-location tasks, subjects should positively predict that agents will search for hidden objects based on their beliefs in both knowledge and FB conditions, but should make no clear predictions about the specific location where an agent will search in ignorance conditions (i.e., when the agent never sees the object being hidden) because the agent has no relevant information to guide goal-directed actions. Strong evidence of belief representation should therefore show that subjects make dissociable predictions depending on whether agents have true, false, or no beliefs. Thus, the inclusion of knowledge and ignorance conditions greatly strengthens our ability to interpret findings from FB conditions.

For example, in [28], apes preferentially helped an agent to open a box (box B) containing an object that he falsely believed to be elsewhere (in box A) when the agent demonstrated a desire to obtain the object by attempting to open box A (*cf* [63,88–90] for replication issues with this task in humans). However, apes in a knowledge condition showed no preference for helping the agent to open the box that the agent acted on, whereas 18-month-old children did show this preference [91]. In addition, apes performed similarly in the knowledge condition and an ignorance condition in which the agent never saw where the object was hidden, suggesting that apes' behavior may not

depend on belief representation. Whereas the original study with 18-month-old children lacked an ignorance condition, in a clever follow-up by Prewasser and colleagues [92] the agent attempted to open a third irrelevant box (box C) which was added to the scene. Although the agent's action here could not have been guided by his FB because he was attempting to open box C (which never contained the object) rather than box A (which he falsely believed the object to be in), children showed similar behavior as in the original FB condition by preferentially helping to open box B (which contained the object). This result suggests that the helping behavior elicited in this task does not depend on the agent's belief. In this case, the inclusion of knowledge and ignorance conditions illuminated theoretical issues with the original interpretation of apes' and childrens' behavior in FB conditions.

In contrast to expectancy violation studies (e.g., [93,94]), neither human [31–54,56–61] nor comparative [25–27] AL studies have included conditions where an agent is completely ignorant about an object's location ([32] for an alternative ignorance control). Kano and colleagues [26] did compare the performance of apes in a FB versus a knowledge condition. However, in that study the target object was always removed from the scene before the test such that the agent did not have knowledge about the specific location of the object, but rather only knowledge that the object was no longer present. Unfortunately, removing the object in this way precludes clear predictions regarding where apes should look in this knowledge condition. Hayashi and colleagues [27] suggested evidence of implicit TB representation based on their familiarization trials in which the agent was knowledgeable, but these trials were not analogous to the FB conditions of the study because the object was never removed (and subjects learned knowledge-congruent looking patterns only after repeated trials). Southgate and colleagues [54] argued that removing the object controls for the possibility that subjects expect agents who lack complete information to search incorrectly in FB conditions, and also for attribution of correct looking in knowledge conditions to a **reality bias** rather than to representation of the agent's knowledge. However, removing the object in this way makes it impossible to establish knowledge or TB conditions in which the subjects are expected to make positive action predictions about a specific location. Underscoring the importance of knowledge conditions with clear associated predictions, in the second of two studies arguing for hints of explicit FB representation in apes [29], four apes passed a FB condition but failed the corresponding knowledge condition. This result makes it unlikely that belief representation guided performance in the FB condition. Therefore, to convincingly demonstrate belief representation, future work should show positive action predictions about agents with knowledge versus FBs that are dissociable from performance in ignorance control conditions.

Implicit but Not Explicit Belief Representation: An Evolutionary Paradox?

If these new comparative studies are indicative of implicit belief representation, the findings raise puzzling evolutionary questions when juxtaposed with numerous convincing failures on explicit FB tasks [18–22]. In humans, implicit understanding of others' beliefs developmentally precedes explicit understanding, and the latter may then be used to guide intentional behavior [5,7,9–12,95]. Although questions remain about the cognitive mechanisms that underpin these representations – including whether they stem from two distinct systems [80,95] or one core system that is later built upon in development [10,11,96,97] – it is clear that explicit belief representation eventually emerges in human development. Explicit belief representation in humans allows flexible prediction of others' actions across diverse contexts [95], a skill which could have clear adaptive benefits (e.g., by enabling intentional deception [98–101] or more flexible cooperation [102,119]).

If primates do have an implicit understanding of beliefs that they cannot explicitly act upon, how did such an understanding evolve? A cognitive system for implicitly tracking FBs that does not guide fitness-relevant behavior would remain invisible to direct evolutionary selective pressure.

Although it remains possible that an implicit understanding of others' beliefs impacts behavior in ways that remain to be discovered, attentional biases produced by belief-congruent looking patterns must ultimately affect behavior for direct selection to occur. Of course, despite many arguments for its potential adaptive value in hominins [103–106], it is also possible that belief representation did not evolve as an adaptation [107–109], but rather as a byproduct of changes in other cognitive systems or via genetic linkage to other fitness-relevant traits [110]. Nonetheless, there are currently no convincing explanations, adaptive or otherwise, for the evolution of implicit but not explicit belief representation, presenting an evolutionary puzzle that must be considered in future work.

Concluding Remarks

Decades after the first comparative studies of ToM [1], it is clear we still have much to learn about the evolution and development of cognitive mechanisms for representing others' beliefs. Although we agree that AL via eye-tracking presents a potentially powerful method for studying belief representation in nonverbal subjects, it is important to recognize both the nascent stage of this research and the unique challenges it presents (see [Outstanding Questions](#)). Given the inherent difficulties of replication studies in primate populations [111] (e.g., few research groups, limited access to subjects, small sample sizes, etc.), it is particularly important to develop and employ consistent and reliable measures (see for instance the ManyPrimates project [112], inspired by ManyBabies [113]). Although we greatly admire recent innovations in this area, we believe that the methodological and theoretical challenges raised here suggest that caution is warranted before concluding that primates can represent others' beliefs.

Acknowledgments

We are grateful to Alyssa Arre for assistance in planning and developing earlier versions of this manuscript. We also thank the members of the Arizona Canine Cognition Center and Yale University Comparative Cognition Laboratory for fruitful discussions on the topic, as well as Christopher Krupenye, Fumihiro Kano, and two anonymous reviewers for their helpful feedback on earlier versions of the manuscript.

Supplemental Information

Supplemental information associated with this article can be found online at <https://doi.org/10.1016/j.tics.2020.05.009>.

References

- Premack, D. and Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 4, 515–526
- Dennett, D.C. (1978) Beliefs about beliefs. *Behav. Brain Sci.* 1, 568–570
- Bennett, J. (1978) Some remarks about concepts. *Behav. Brain Sci.* 1, 557–560
- Pylyshyn, Z.W. (1978) When is attribution of beliefs justified? *Behav. Brain Sci.* 1, 592–592
- Wimmer, H. and Perner, J. (1983) Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128
- Baron-Cohen, S. *et al.* (1985) Does the autistic child have a 'theory of mind'? *Cognition* 21, 37–46
- Wellman, H.M. *et al.* (2001) Meta-analysis of theory of mind development: the truth about false belief. *Child Dev.* 72, 655–684
- Wellman, H.M. and Liu, D. (2004) Scaling of theory-of-mind tasks. *Child Dev.* 75, 523–541
- Wellman, H.M. (2018) Theory of mind: the state of the art. *Eur. J. Dev. Psychol.* 15, 728–755
- Kovacs, A.M. *et al.* (2010) The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834
- Onishi, K.H. and Baillargeon, R. (2005) Do 15-month-old infants understand false beliefs? *Science* 308, 255–258
- Baillargeon, R. *et al.* (2010) False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118
- Scott, R.M. and Baillargeon, R. (2017) Early false-belief understanding. *Trends Cogn. Sci.* 21, 237–249
- Call, J. and Tomasello, M. (2008) Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192
- Rosati, A.G. *et al.* (2010) Primate social cognition: thirty years after Premack and Woodruff. In *Primate Neuroethology* (Platt, M. and Ghazanfar, A., eds), pp. 117–144, Oxford University Press
- Krupenye, C. and Call, J. (2019) Theory of mind in animals: current and future directions. *Wiley Interdiscip. Rev. Cogn. Sci.* 10, e1503
- Arre, A.M. and Santos, L.R. Mentalizing in non-human primates. In *Neural Basis of Mentalizing* (Ochsner, K. and Gilead, M., eds), Springer, in press
- Call, J. and Tomasello, M. (1999) A nonverbal false belief task: the performance of children and great apes. *Child Dev.* 70, 381–395
- Kaminski, J. *et al.* (2008) Chimpanzees know what others know, but not what they believe. *Cognition* 109, 224–23
- Krachun, C. *et al.* (2009) A competitive nonverbal false belief task for children and apes. *Dev. Sci.* 12, 521–535
- Krachun, C. *et al.* (2010) A new change-of-contents false belief test: children and chimpanzees compared. *Int. J. Comp. Psychol.* 23, 145–165
- Hare, B. *et al.* (2001) Do chimpanzees know what conspecifics know? *Anim. Behav.* 61, 139–151

Outstanding Questions

Numerous strict replication attempts of AL belief representation tests using the original stimuli and procedures have recently failed to reproduce original results in the human literature. Why are ToM tasks using AL less replicable than those using other measures? What procedural variations account for conceptual non-replications?

None of the three recent studies reporting evidence of belief representation in non-human primates using AL tasks included a condition where the agent was knowledgeable about the specific location of an object or was completely ignorant about an object's location. Would non-human primates show positive action predictions about agents with knowledge versus FBs that are dissociable from performance in ignorance control conditions?

No previous work has tested belief representation in apes using expectancy violation measures despite theoretical arguments that post-event epistemic-state tracking may be less cognitively demanding than the online tracking demanded by AL. How would apes perform on expectancy violation FB tests?

In humans, an implicit understanding of others' beliefs develops before explicit representation which is then used to guide intentional behavior. Given that natural selection acts on behavior, why and how might a strictly implicit understanding of others' beliefs have evolved?

23. Marticorena, D.C.W. *et al.* (2011) Monkeys represent others' knowledge but not their beliefs. *Dev. Sci.* 4, 1406–1416
24. Martin, A. and Santos, L.R. (2014) The origins of belief representation: monkeys fail to automatically represent others' beliefs. *Cognition* 130, 300–308
25. Krupenye, C. *et al.* (2016) Great apes anticipate that other individuals will act according to false beliefs. *Science* 354, 110–113
26. Kano, F. *et al.* (2019) Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proc. Natl. Acad. Sci. U. S. A.* 116, 20904–20909
27. Hayashi, T. *et al.* (2020) Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Rep.* 30, 4433–4444
28. Buttelmann, D. *et al.* (2017) Great apes distinguish true from false beliefs in an interactive helping task. *PLoS One* 12, e0173793
29. O'Connell, S. and Dunbar, R. (2003) A test for comprehension of false belief in chimpanzees. *Evol. Cogn.* 9, 131–139
30. Joiner, J.A. (2020) *Social Information Processing in Rhesus Monkeys*, Yale University Press
31. Clements, W.A. and Perner, J. (1994) Implicit understanding of belief. *Cogn. Dev.* 9, 377–395
32. Garnham, W.A. and Ruffman, T. (2001) Doesn't see, doesn't know: is anticipatory looking really related to understanding of belief? *Dev. Sci.* 4, 94–100
33. Schneider, D. *et al.* (2012) Cognitive load disrupts implicit theory-of-mind processing. *Psychol. Sci.* 23, 842–847
34. Surian, L. and Geraci, A. (2012) Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *Br. J. Dev. Psychol.* 30, 30–44
35. Thoermer, C. *et al.* (2012) Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *Br. J. Dev. Psychol.* 30, 172–187
36. Low, J. and Watts, J. (2013) Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychol. Sci.* 24, 305–311
37. Gilga, T. *et al.* (2014) Spontaneous belief attribution in younger siblings of children on the autism spectrum. *Dev. Psychol.* 50, 903–913
38. Schuwerk, T. *et al.* (2015) Implicit and explicit theory of mind reasoning in autism spectrum disorders: the impact of experience. *Autism* 19, 459–468
39. Wang, B. *et al.* (2015) Limits on efficient human mindreading: convergence across Chinese adults and Semai children. *Br. J. Psychol.* 106, 724–740
40. Zmyj, N. *et al.* (2015) Eighteen-month-olds' memory interference and distraction in a modified A-not-B task is not associated with their anticipatory looking in a false-belief task. *Front. Psychol.* 6, 857
41. Meristo, M. *et al.* (2016) Early conversational environment enables spontaneous belief attribution in deaf children. *Cognition* 157, 139–145
42. Schuwerk, T. *et al.* (2016) Implicit mentalizing persists beyond early childhood and is profoundly impaired in children with autism spectrum condition. *Front. Psychol.* 7, 1696
43. Ruffman, T. *et al.* (2001) Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *J. Exp. Child Psychol.* 80, 201–224
44. Sodian, B. *et al.* (2016) Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: a longitudinal investigation. *Child Dev.* 87, 1221–1232
45. Wang, L. and Leslie, A.M. (2016) Is implicit theory of mind the 'real deal'? The own-belief/true-belief default in adults and young preschoolers. *Mind Lang.* 31, 147–176
46. Grosse Wiesmann, C. *et al.* (2017) Implicit and explicit false belief development in preschool children. *Dev. Sci.* 20, e12445
47. Burnside, K. *et al.* (2018) Implicit false belief across the lifespan: non-replication of an anticipatory looking task. *Cogn. Dev.* 46, 4–11
48. Dörrenberg, S. *et al.* (2018) How (not) to measure infant theory of mind: testing the replicability and validity of four non-verbal measures. *Cogn. Dev.* 46, 12–30
49. Grosse Wiesmann, C. *et al.* (2018) Longitudinal evidence for 4-year-olds' but not 2- and 3-year-olds' false belief-related action anticipation. *Cogn. Dev.* 46, 58–68
50. Kulke, L. *et al.* (2018) How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cogn. Dev.* 46, 97–111
51. Kulke, L. *et al.* (2018) Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychol. Sci.* 29, 888–900
52. Schuwerk, T. *et al.* (2018) The robustness and generalizability of findings on spontaneous false belief sensitivity: a replication attempt. *R. Soc. Open Sci.* 5, 172273
53. Kulke, L. *et al.* (2019) Is implicit theory of mind real but hard to detect? Testing adults with different stimulus materials. *R. Soc. Open Sci.* 6, 190068
54. Southgate, V. *et al.* (2007) Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18, 587–592
55. Grosse Wiesmann, C. *et al.* (2020) Two systems for thinking about others' thoughts in the developing brain. *Proc. Natl. Acad. Sci. U. S. A.* 117, 6928–6935
56. Senju, A. *et al.* (2009) Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science* 325, 883–885
57. Senju, A. *et al.* (2010) Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Dev. Psychopathol.* 22, 353–360
58. Low, J. (2010) Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Dev.* 81, 597–615
59. Senju, A. *et al.* (2011) Do 18-month-olds really attribute mental states to others?: a critical test. *Psychol. Sci.* 22, 878–880
60. Meristo, M. *et al.* (2012) Belief attribution in deaf and hearing infants. *Dev. Sci.* 15, 633–640
61. Schneider, D. *et al.* (2012) Eye movements reveal sustained implicit processing of others' mental states. *J. Exp. Psychol. Gen.* 141, 433–438
62. Baillargeon, R. *et al.* (2018) Invited commentary. Interpreting failed replications of early false-belief findings: methodological and theoretical considerations. *Cogn. Dev.* 46, 112–124
63. Poulin-Dubois, D. *et al.* (2018) Do infants understand false beliefs? We don't know yet – a commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cogn. Dev.* 48, 302–315
64. Kulke, L. and Rakoczy, H. (2018) Implicit theory of mind – an overview of current replications and non-replications. *Data Brief* 16, 101–104
65. Barone, P. *et al.* (2019) Infants' performance in spontaneous-response false belief tasks: a review and meta-analysis. *Infant Behav. Dev.* 57, 101350
66. Powell, L.J. *et al.* (2018) Replications of implicit theory of mind tasks with varying representational demands. *Cogn. Dev.* 46, 40–50
67. Yott, J. and Poulin-Dubois, D. (2016) Are infants' theory-of-mind abilities well integrated? Implicit Understanding of intentions, desires, and beliefs. *J. Cogn. Dev.* 17, 683–698
68. Rubio-Fernández, P. (2019) Publication standards in infancy research: three ways to make violation-of-expectation studies more reliable. *Infant Behav. Dev.* 54, 177–188
69. Krupenye, C. *et al.* (2017) A test of the submentalizing hypothesis: apes' performance in a false belief task inanimate control. *Commun. Integr. Biol.* 10, e1343771
70. Völter, C.J. *et al.* (2019) Chimpanzees flexibly update working memory contents and show susceptibility to distraction in the self-ordered search task. *Proc. R. Soc. B Biol. Sci.* 286, 20190715
71. Rosati, A.G. (2017) Chimpanzee cognition and the roots of the human mind. In *Chimpanzees and Human Evolution* (Muller, M.N. *et al.*, eds), pp. 703–745, Harvard University Press
72. Rosati, A.G. (2017) The evolution of primate executive function: from response control to strategic decision-making. In *Evolution of Nervous Systems* (Vol. 3) (2nd edn) (Kaas, J. and Krubitzer, L., eds), pp. 423–437, Elsevier
73. Beran, M.J. and Evans, T.A. (2006) Maintenance of delay of gratification by four chimpanzees (*Pan troglodytes*): the effects of delayed reward visibility, experimenter presence, and extended delay intervals. *Behav. Process.* 73, 315–324
74. Evans, T.A. and Beran, M.J. (2007) Chimpanzees use selfdistraction to cope with impulsivity. *Biol. Lett.* 3, 599–602

75. Warneken, F. and Rosati, A.G. (2015) Cognitive capacities for cooking in chimpanzees. *Proc. R. Soc. B Biol. Sci.* 282, 20150229
76. Herrmann, E. *et al.* (2015) Uniquely human self-control begins at school age. *Dev. Sci.* 18, 979–993
77. Rosati, A.G. *et al.* (2007) The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. *Curr. Biol.* 17, 1663–1668
78. Penn, D.C. and Povinelli, D.J. (2007) On the lack of evidence that non-human animals possess anything remotely resembling a ‘theory of mind’. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 362, 731–744
79. Heyes, C. (2014) Submentalizing: I am not really reading your mind. *Perspect. Psychol. Sci.* 9, 131–143
80. Butterfill, S.A. and Apperly, I.A. (2013) How to construct a minimal theory of mind. *Mind Lang.* 28, 606–637
81. Perner, J. and Roessler, J. (2012) From infants’ to children’s appreciation of belief. *Trends Cogn. Sci.* 16, 519–525
82. Martin, A. and Santos, L.R. (2016) What cognitive representations support primate theory of mind? *Trends Cogn. Sci.* 20, 375–382
83. Phillips, J. and Norby, A. (2019) Factive theory of mind. *Mind Lang.* Published online November 27, 2019. <https://doi.org/10.1111/mla.12267>
84. Tomasello, M. (2018) How children come to understand false beliefs: a shared intentionality account. *Proc. Natl. Acad. Sci. U. S. A.* 115, 8491–8498
85. Gettier, E.L. (1963) Is justified true belief knowledge? *Analysis* 23, 121–123
86. Horschler, D.J. *et al.* (2019) Do non-human primates really represent others’ ignorance? A test of the awareness relations hypothesis. *Cognition* 190, 72–80
87. Ruffman, T. (1996) Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind Lang.* 11, 388–414
88. Crivello, C. and Poulin-Dubois, D. (2018) Infants’ false belief understanding: a non-replication of the helping task. *Cogn. Dev.* 46, 51–57
89. Fizke, E. *et al.* (2017) Are there signature limits in early theory of mind? *J. Exp. Child Psychol.* 162, 209–224
90. Oktay-Gür, N. *et al.* (2018) Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition* 173, 60–74
91. Buttelmann, D. *et al.* (2009) Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112, 337–342
92. Prieuasser, B. *et al.* (2018) Helping as an early indicator of a theory of mind: mentalism or teleology? *Cogn. Dev.* 46, 69–78
93. He, Z. *et al.* (2011) False-belief understanding in 2.5-year-olds: evidence from violation-of-expectation change-of-location and unexpected-contents tasks. *Dev. Sci.* 14, 292–305
94. Scott, R.M. and Baillargeon, R. (2009) Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Dev.* 80, 1172–1196
95. Apperly, I.A. and Butterfill, S.A. (2009) Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970
96. Leslie, A.M. (1994) ToMM, ToBy, and Agency: core architecture and domain specificity. In *Mapping the Mind: Domain Specificity in Cognition and Culture* (Hirschfeld, L.A. and Gelman, S.A., eds), pp. 119–143, Cambridge University Press
97. Leslie, A.M. *et al.* (2004) Core mechanisms in ‘theory of mind’. *Trends Cogn. Sci.* 8, 528–533
98. Byrne, R.W. and Whiten, A. (1996) Machiavellian intelligence. *Evol. Anthropol.* 5, 172–180
99. Whiten, A. and Byrne, R.W. (1988) Tactical deception in primates. *Behav. Brain Sci.* 11, 233–273
100. Hare, B. *et al.* (2006) Chimpanzees deceive a human competitor by hiding. *Cognition* 101, 495–514
101. Kummer, H. *et al.* (1997) The social intelligence hypothesis. In *Human by Nature: Between Biology and the Social Sciences* (Weingart, P. *et al.*, eds), pp. 157–179, Lawrence Erlbaum Associates
102. Tomasello, M. *et al.* (2012) Two key steps in the evolution of human cooperation. *Curr. Anthropol.* 53, 673–692
103. Fletcher, L. and Carruthers, P. (2013) Behavior-reading versus mentalizing in animals. In *Agency and Joint Attention* (Metcalfe, J. and Terrace, H., eds), pp. 82–99, Oxford University Press
104. Corballis, M.C. (2007) The uniqueness of human recursive thinking. *Am. Sci.* 95, 240–248
105. Tomasello, M. (2008) *Origins of Human Communication*, MIT Press
106. Richerson, P.J. and Boyd, R. (2008) *Not by Genes Alone: How Culture Transformed Human Evolution*, University of Chicago Press
107. Gould, S.J. and Lewontin, R.C. (1979) The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 205, 581–598
108. Gould, S.J. and Vrba, E.S. (1982) Exaptation – a missing term in the science of form. *Paleobiology* 8, 4–15
109. Bateson, P. and Laland, K.N. (2013) Tinbergen’s four questions: an appreciation and an update. *Trends Ecol. Evol.* 28, 712–718
110. Lloyd, E.A. (2015) Adaptationism and the logic of research questions: how to think clearly about evolutionary causes. *Biol. Theory* 10, 343–362
111. Stevens, J.R. (2017) Replicability and reproducibility in comparative psychology. *Front. Psychol.* 8, 862
112. Altschul, D.M. *et al.* (2019) Establishing an infrastructure for collaboration in primate cognition research. *PLoS One* 14, e0223675
113. Frank, M.C. *et al.* (2017) A collaborative approach to infant research: promoting reproducibility, best practices, and theory-building. *Infancy* 22, 421–435
114. Meltzoff, A.N. and Brooks, R. (2008) Self-experience as a mechanism for learning about others: a training study in social cognition. *Dev. Psychol.* 44, 1257–1265
115. Karg, K. *et al.* (2015) The goggles experiment: can chimpanzees use self-experience to infer what a competitor can see? *Anim. Behav.* 105, 211–221
116. Povinelli, D.J. and Vonk, J. (2003) Chimpanzee minds: suspiciously human? *Trends Cogn. Sci.* 7, 157–160
117. Heyes, C. (1998) Theory of mind in nonhuman primates. *Behav. Brain Sci.* 21, 101–148
118. Oktay-Gür, N. and Rakoczy, H. (2017) Children’s difficulty with true belief tasks: competence deficit or performance problem? *Cognition* 166, 28–41
119. Baillargeon, R. *et al.* (2013) False-belief understanding and why it matters. Navigating the social world: what infants, children, and other species can teach us. In *In Navigating the Social World*, pp. 88–95, Oxford University Press