Title: Great apes anticipate that other individuals will act according to false beliefs

Authors: Christopher Krupenye[1]†*, Fumihiro Kano[2,3]†*, Satoshi Hirata[2], Josep Call[4,5], and Michael Tomasello[5,6]

Affiliations:

[1]Department of Evolutionary Anthropology, Duke University, Durham, USA

[2]Kumamoto Sanctuary, Wildlife Research Center, Kyoto University, Kumamoto, Japan

[3]Primate Research Institute, Kyoto University, Inuyama, Japan

[4]School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, UK

[5]Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[6]Department of Psychology and Neuroscience, Duke University, Durham, USA

†Shared first-authorship

*Corresponding authors: ckrupenye@gmail.com; fkanou@gmail.com

Abstract:

Humans operate with a "theory-of-mind" with which they understand that others' actions are driven not by reality but by beliefs about reality, even when those beliefs are false. Although great apes share with humans many social-cognitive skills, they have repeatedly failed experimental tests of such false belief understanding. Using an anticipatory looking test (originally developed for human infants), we show that three species of great apes reliably look in anticipation of an agent acting on a location where he falsely believes an object to be, even though they themselves know that it is no longer there. These results suggest that great apes also operate—at least on an implicit level—with an understanding of false beliefs.

One Sentence Summary: Great apes appear to understand false beliefs.

Main Text:

Central to everything that makes us human—including our unique modes of communication, cooperation, and culture—is our *theory of mind* (ToM). ToM is the ability to impute unobservable mental states, such as desires and beliefs, to others (*1, 2*). For nearly four decades, a cardinal question in psychology has concerned whether nonhuman animals, such as great apes, also possess this cognitive skill (*1, 3*). A variety of non-verbal behavioral experiments have provided converging evidence that apes can predict others' behavior not simply based on external cues but rather on an understanding of their goals, intentions, and knowledge (*3, 4*). However, it remains unclear whether apes can comprehend reality-incongruent mental states, e.g., false beliefs (*3*), and indeed apes have failed to make explicit behavioral choices that reflect false belief understanding in several food-choice tasks (*4-6*). False belief understanding is of particular interest because it requires understanding that others' actions are driven not by reality but by beliefs about reality, even when those beliefs are false.

In human developmental studies, it is only after age four that children pass traditional false belief tests, in which they must explicitly predict a mistaken agent's future actions (*7*). However, recent evidence has shown that even young infants can pass modified false belief tests using simplified task procedures and spontaneous-gaze responses as measures (e.g., violation-of-expectation (*8*); anticipatory-looking (*9, 10*)). For example, anticipatory looking paradigms exploit individuals' tendency to look to a location in anticipation of an impending event, and thus can measure a participant's predictions about what an agent is about to do—even when that agent holds a false belief about the situation. Only two studies have used spontaneous-gaze false belief tasks with nonhuman primates. Both failed to replicate with monkeys the results with infants, despite monkeys' success in true-belief conditions (*11, 12*).

Here we use an anticipatory-looking measure (*10*) to test for false belief understanding in three species of apes (chimpanzees, *Pan troglodytes*; bonobos, *Pan paniscus*; orangutans, *Pongo abelii*). Previous studies have established that apes reliably make anticipatory looks based on agents' goal-directed actions and subjects' event memories (*13, 14*). In our experiments, apes watched short videos on a monitor while their gaze was noninvasively recorded using an infrared eye-tracker. Our design, controls, and general procedure replicated a seminal anticipatory-looking false belief study with human infants (*10*).

We conducted a pair of experiments using the same design but introduced unique scenarios in each. The common design involved two familiarization trials followed by a single test trial (either the FB1 or FB2 condition; between-subjects design). In our scenarios, a human agent pursued a goal object that was hidden in one of two locations. During the first familiarization, the agent witnessed the hiding of the object in one location before searching for it there. In the second, the object was hidden in the other location and the agent pursued it there. These trials served to demonstrate that the object could be hidden in either of the two locations and that, when knowledgeable, the agent would search for it in its true location. During the belief-induction phase, the agent witnessed the initial hiding of the object, but the object was then moved to a second location, while the agent was either present (FB1) or absent (FB2). In both conditions, the object was then completely removed before the agent returned to search for it. The agent's actions presented during the induction phase controlled for several low-level cues, namely that participants could not solve the task by simply expecting the agent to search in the first or last location where the object was hidden or the last location where the agent attended (*10*). Whether the object was hidden first in the left or right location during familiarization trials and whether the Target of the agent's false belief was the left or right location during test trials was counterbalanced across subjects.

Experiments 1 and 2 presented unique scenarios intended to evoke apes' spontaneous action anticipation in different contexts. To encourage subjects' engagement, we presented simulated agonistic encounters between a human (Actor) and King-Kong (KK), an unreal ape-like character unfamiliar to the subjects (*14*). To minimize the possibility that apes could solve the task by responding to learned behavioral cues, our scenarios involved events that were novel to our participants. In Experiment 1, the Actor attempted to search for KK, who had hidden himself in one of two large haystacks (Figure 1 and Movie S1). In Experiment 2, the Actor attempted to retrieve a stone that KK had stolen and hidden in one of two boxes (Figure 2 and Movie S2). We confirmed that apes unambiguously attended to the depicted actions during the belief-induction phases of both experiments (Figures S3 and S4; *15*).

Apes' anticipatory looks were assessed based on their first looks to the Target (the location where the Actor falsely believed the object to be) or Distractor (the other location) as the Actor ambiguously approached the two locations—from the start to the end of the Actor's walk toward the haystacks (*central-approach*; Experiment 1; Figures 1k and 1p) and reach toward the boxes (*central-reach*; Experiment 2; Figures 2n and 2t) (both 4.5 seconds). Software scored looks automatically based on Areas-Of-Interest (Figures 1q and 2u). Importantly, the Actor's gaze and gait during the central-approach and central-reach provided no directional cues (Figures S1 and S2; *15*) and the videos ended without the Actor hitting or grabbing the Target. We used two different scenarios to gauge the robustness of apes' responses under different conditions.

Table 1 summarizes the results for each experiment. In Experiment 1, we tested 40 apes (19 chimpanzees, 14 bonobos, and 7 orangutans; Table S1; *15*). Thirty subjects looked to either the Target or Distractor during the central-approach period. Of 30, 20 looked first at the Target (p = 0.098, two-tailed binomial test). There was no difference between FB1 and FB2 conditions (p = 0.70, Fisher's exact test). In Experiment 2, we tested 30 subjects (29 from Experiment 1 plus one additional bonobo). Twenty-two apes made explicit looks to the Target or Distractor during this period. Of 22, 17 looked first at the Target (p = 0.016, two-tailed binomial test), and there was no difference between FB1 and FB2 conditions (p = 1.0, Fisher's exact test).

Critically, we conducted a combined analysis with the 29 apes who participated in both experiments. We compared the number of first looks each subject made to the Target versus to the Distractor during the central-approach and central-reach periods (i.e., max 2; Figure 3). Apes made significantly more first looks to the Target than to the Distractor, overall (Wilcoxon signed rank test: Z = 3.25, N = 29, p = 0.001, r = 0.42), and also in each condition (FB1: Z = 1.98, N = 15, p = 0.046, r = 0.36; FB2: Z = 2.15, N = 14, p = 0.031, r = 0.40; Figure 3A). No significant difference was detected across species. To test this, we first calculated difference scores for each subject (number of first looks to Target minus to Distractor) and then subjected them to the Kruskal-Wallis H test ($\chi^2(2)$ = 0.46, p = 0.79, Figure 3B).

Our findings show that apes accurately anticipated the goal-directed behavior of an agent who held a false belief. Our design and results controlled for several alternatives. First, apes could not solve the task by simply expecting the Actor to search in the first or last location where the object was hidden, the last location the Actor attended, or the last location KK acted on. Second, apes could not merely *respond* to violations of three-way associations between the Actor, the target object, and the object's location, formed during familiarization or belief-induction phases (*16*); instead they *actively predicted* the Actor's behavior. Heyes (*17*) argued that a low-level account could explain Southgate et al.'s (*10*) results if subjects overlooked the object's movement while the agent was not attending, and imagined the object in its previous location. Critically, we confirmed that apes closely tracked all such movements (Figures S3-S4; *15*). Third, our results cannot be explained as attribution of ignorance rather than false belief. Apes did not simply expect the Actor's ignorance to lead to error or uncertainty (*18*); they specifically anticipated that the Actor would search for the object where he falsely believed it to be.

Apes were never shown the Actor's search behavior when he held a false belief, precluding reliance on external behavioral cues learned during the task. We also minimized the possibility that subjects could apply behavior rules that they have acquired through extensive learning during past experiences, by requiring subjects to make predictions in situations that involved a constellation of novel features (e.g., a human attacking an ape-like character hiding in a haystack). However, we acknowledge that all change-of-location false belief tasks are in principle open to an abstract behavior rule-based explanation, namely, that apes could solve the task by relying on a rule that agents search for things where they last saw them (*16*). Importantly, however, this explanatory framework cannot easily accommodate the diversity of existing evidence for ape ToM (*3*), or recent evidence that human infants and apes appear to infer whether others can see through objects that *look* opaque based on their own self-experience with the occlusive properties (i.e., see-through or opaque) of those objects (*19, 20*).

Our results in concert with existing data thus suggest that apes solved the task by ascribing a false belief to the Actor, challenging the view that the ability to attribute reality-incongruent mental states is unique to humans. Given that apes have not yet succeeded on tasks that measure false belief understanding based on explicit behavioral choices (*4-6*), the present evidence may constitute an implicit understanding of belief (*9*). Differential performance may reflect differences in task demands or context, or less flexible abilities in apes compared to humans. At minimum, apes can anticipate that an actor will pursue a goal object where he last saw it, even though they themselves know that it is no longer there. That great apes operate—at least on an implicit level—with an understanding of false beliefs suggests that this essential ToM skill is likely at least as old as humans' last common ancestor with the other apes.

References and Notes:

1. D. Premack, G. Woodruff, Does the chimpanzee have a theory of mind. *Behav. Brain Sci.* **1**, 515-526 (1978).
2. R. W. Byrne, A. W. Whiten, *Machiavellian intelligence : social expertise and the evolution of intellect in monkeys, apes, and humans*. (Clarendon Press, Oxford, 1988).
3. J. Call, M. Tomasello, Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* **12**, 187-192 (2008).
4. B. Hare, J. Call, M. Tomasello, Do chimpanzees know what conspecifics know? *Animal Behaviour* **61**, 139-151 (2001).
5. J. Kaminski, J. Call, M. Tomasello, Chimpanzees know what others know, but not what they believe. *Cognition* **109**, 224-234 (2008).
6. C. Krachun, M. Carpenter, J. Call, M. Tomasello, A competitive nonverbal false belief task for children and apes. *Dev. Sci.* **12**, 521-535 (2009).
7. H. M. Wellman, D. Cross, J. Watson, Meta-analysis of theory-of-mind development: The truth about false belief. *Child Dev.*, 655-684 (2001).
8. K. H. Onishi, R. Baillargeon, Do 15-month-old infants understand false beliefs? *Science* **308**, 255-258 (2005).
9. W. A. Clements, J. Perner, Implicit understanding of belief. *Cogn. Dev.* **9**, 377-395 (1994).
10. V. Southgate, A. Senju, G. Csibra, Action anticipation through attribution of false belief by 2-year-olds. *Psychol Sci* **18**, 587-592 (2007).

11. D. C. Marticorena, A. M. Ruiz, C. Mukerji, A. Goddu, L. R. Santos, Monkeys represent others' knowledge but not their beliefs. *Dev. Sci.* **14**, 1406-1416 (2011).

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

13. F. Kano, J. Call, Great apes generate goal-based action predictions: an eye-tracking study. *Psychol Sci* **25**, 1691-1698 (2014).

14. F. Kano, S. Hirata, Great apes make anticipatory looks based on long-term memory of single events. *Curr. Biol.* **25**, 2513-2517 (2015).

15. Materials and methods are available as supplementary materials at the Science website.

16. J. Perner, T. Ruffman, Infants' insight into the mind: How deep? *Science* **308**, 214-216 (2005).

17. C. Heyes, False belief in infancy: a fresh look. *Dev Sci* **17**, 647-659 (2014).

18. R. Baillargeon, R. M. Scott, Z. He, False-belief understanding in infants. *Trends Cogn. Sci.* **14**, 110-118 (2010).

19. A. Senju, V. Southgate, C. Snape, M. Leonard, G. Csibra, Do 18-month-olds really attribute mental states to others? A critical test. *Psychol. Sci.* **22**, 878-880 (2011).

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

Acknowledgements:

We thank the staff at the Wolfgang Kohler Primate Research Center and Kumamoto Sanctuary for assistance. Financial support came from NSFGRFP DGE-1106401 (CK), MEXT K-CONNEX, JSPS KAKENHI 26885040, 16K21108 (FK), JSPS KAKENHI 26245069, 24000001 (SH), and ERC-Synergy grant 609819 (JC). Data are available in the main text and supplementary materials.

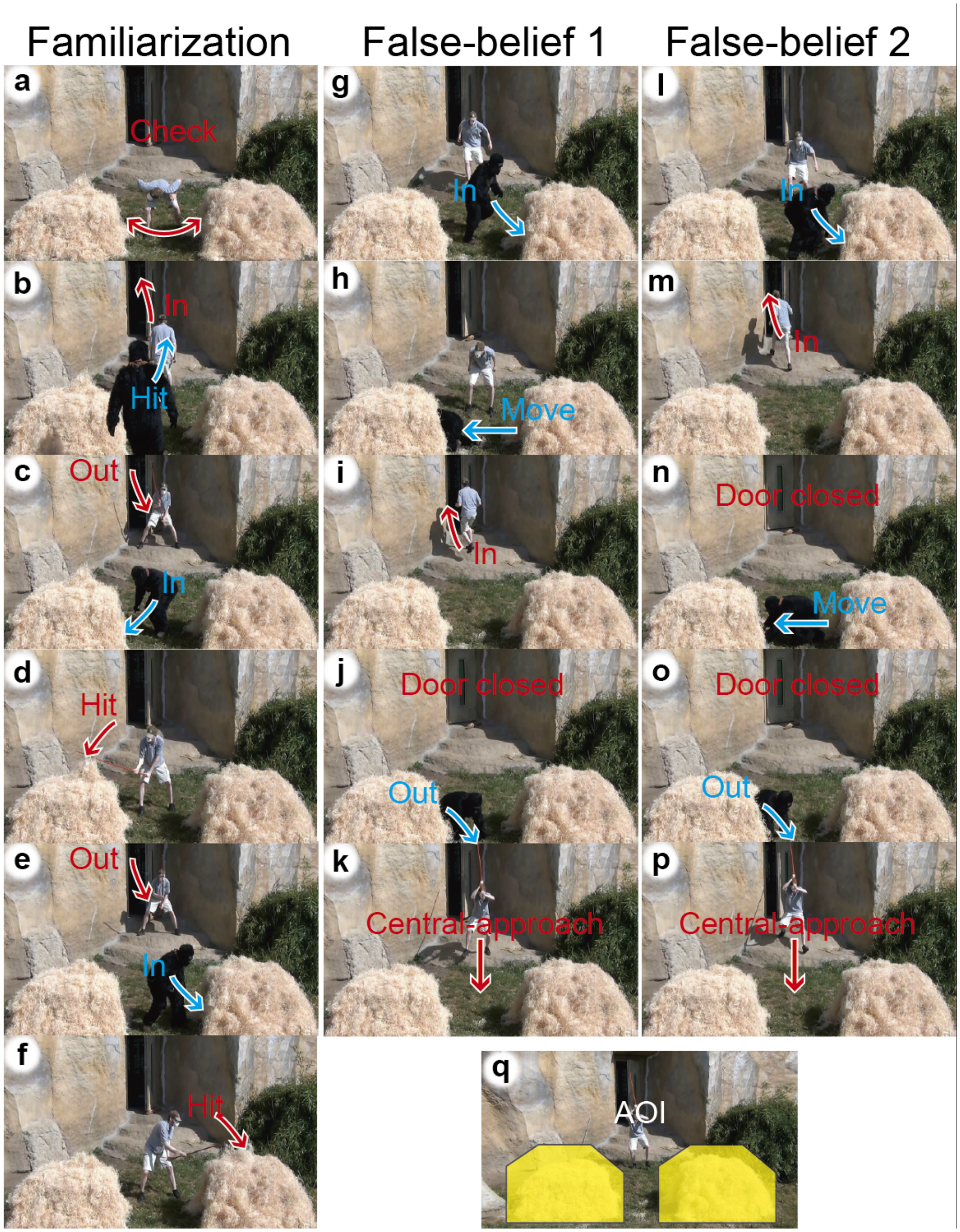

Figure 1. Events shown during familiarization (a-f), belief-induction (g-j in FB1 condition; l-o in FB2 condition), and central-approach (k in FB1; p in FB2) in Experiment 1. Also shown are the Areas-of-Interest (AOIs) defined for the Target and Distractor haystacks (q). See Movie S1.

Figure 2. Events shown during familiarization (a-g), belief-induction (h-m in FB1 condition; o-t in FB2 condition), and central-reach (n in FB1; u in FB2) in Experiment 2.

Following the infant study (*10*) , we included an additional action in FB1 (KK touched the

Distractor box; k) to control for subjects looking to the last place that the actor attended. Also shown are the Areas-of-Interest (AOIs) defined for the Target and Distractor boxes (u). See Movie S2.

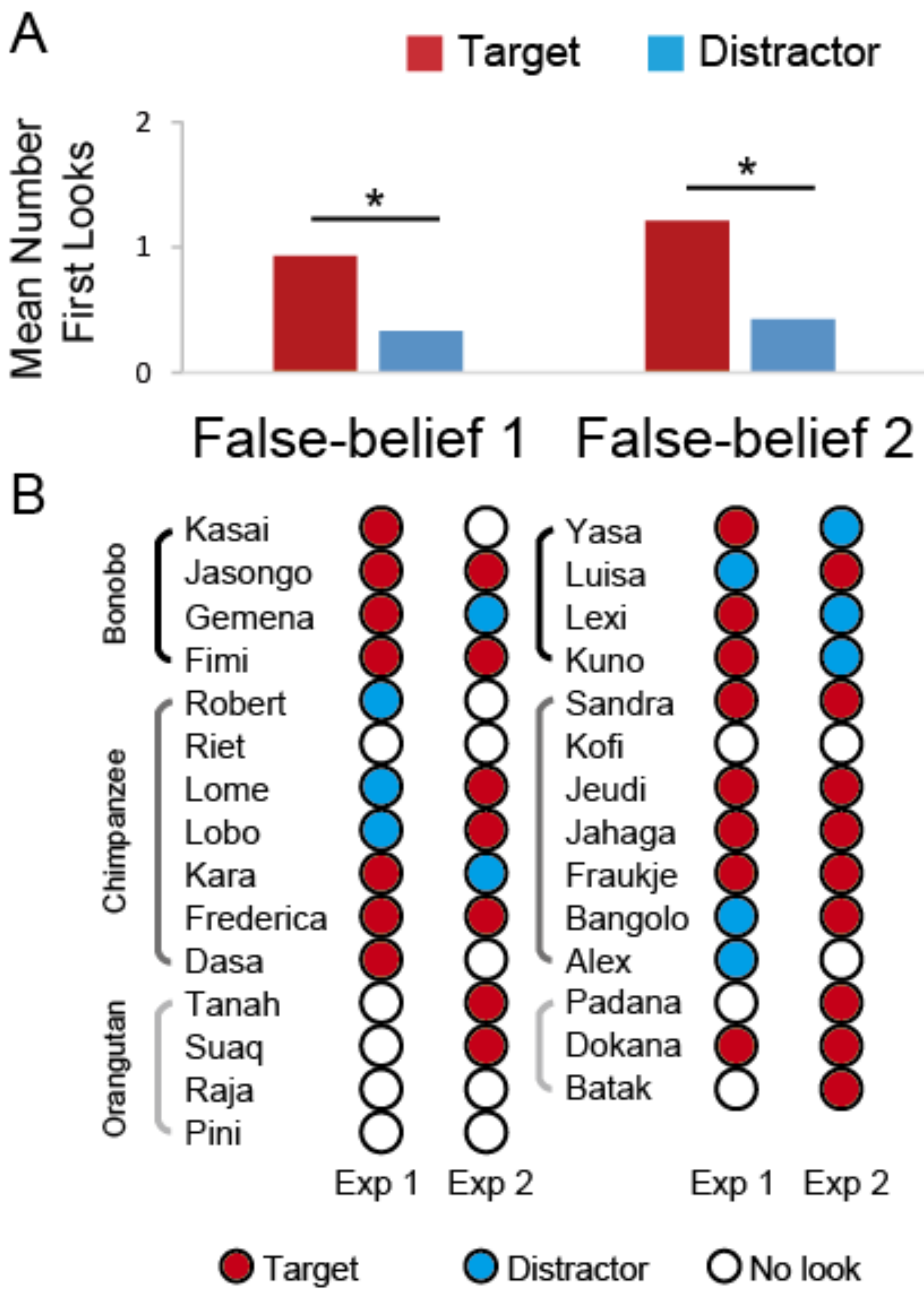

Figure 3. Apes' performance across the two experiments.

A. Mean number of first looks to the Target and to the Distractor for the 29 subjects who participated in *both* Experiments 1 and 2. Asterisks indicate $p<0.05$, Wilcoxon signed rank test.

B. Individual scores in each experiment.

Table 1. Number of participants who made first looks to either the Target or Distractor during the agent's approach in Experiments 1 (*N*=40) and 2 (*N*=30). Shown in parentheses is the number of participants who did not look at either.

		Target	Distractor	Total
Experiment 1	FB1	10	4	14 (6)
	FB2	10	6	16 (4)
	Total	20	10	30 (10)
Experiment 2	FB1	8	2	10 (6)
	FB2	9	3	12 (2)
	Total	17	5	22 (8)

Supplementary Materials:

Materials and Methods

Supporting Text

Figures S1-S5

Tables S1-S5

Captions for Movies S1-S2

Movies S1-S2