

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

Christopher Krupenye^{1, *}, Fumihiro Kano^{2,3, *}, Satoshi Hirata², Josep Call^{4,5}, Michael Tomasello^{5,6}

* shared first-authors, co-correspondence

1. Department of Evolutionary Anthropology, Duke University
2. Kumamoto Sanctuary, Wildlife Research Center, Kyoto University
3. Primate Research Institute, Kyoto University
4. School of Psychology and Neuroscience, University of St Andrews
5. Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology
6. Department of Psychology and Neuroscience, Duke University

Abstract

Humans operate with a “theory of mind” with which they are able to 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. We use an anticipatory looking test (originally developed for human infants) to 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 the apes themselves know that the object is no longer there. Our results suggest that great apes also operate, at least on an implicit level, with an understanding of false beliefs.

Central to everything that makes us human—including our distinctive 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 nonverbal behavioral experiments have provided converging evidence that apes can predict others’ behavior, not simply based on external cues but rather on an understanding of others’ goals, perception, and knowledge (3, 4). However, it remains unclear whether apes can comprehend reality-incongruent mental states (e.g., false beliefs) (3), as 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 recognizing 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 4 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 that involve the use of 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).

In our study, we used 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 distinct scenarios in each. The common design involved two familiarization trials followed by a single test trial [either the FB1 or FB2 (false belief one or two) 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 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 were counterbalanced across subjects.

Experiments one and two presented scenarios that were specifically 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 apelike 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 one, the actor attempted to search for KK, who had hidden himself in one of two large haystacks (Fig. 1 and movie S1). In experiment two, the actor attempted to retrieve a stone that KK had stolen and hidden in one of two boxes (Fig. 2 and movie S2). We confirmed that apes unambiguously attended to the depicted actions during the belief-induction phases of both experiments (figs. S3 and S4) (15).

Apes' anticipatory looks were assessed on the basis of their first looks to the target (the location where the actor falsely believed the object to be) or the 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 one (Fig. 1, K and P)] and reach toward the boxes [central reach; experiment two (Fig. 2, N and T)] (both 4.5 s). Software scored looks automatically on the basis of areas of interest (15) (Figs. 1Q and 2U). The actor's gaze and gait during the central approach and central reach provided no directional cues (figs. 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 one, we tested 40 apes [19 chimpanzees, 14 bonobos, and 7 orangutans (table S1) (15)]. Thirty subjects looked to either the target or the distractor during the central-approach period. Of these 30, 20 looked first at the target ($P = 0.098$, two-tailed binomial test). There was no difference between the FB1 and FB2 conditions ($P = 0.70$, Fisher's exact test). In experiment two, we tested 30 subjects (29 from experiment one,

plus one additional bonobo). Twenty-two apes made explicit looks to the target or the distractor during this period. Of these 22, 17 looked first at the target ($P = 0.016$, two-tailed binomial test), and there was no difference between the FB1 and FB2 conditions ($P = 1.0$, Fisher's exact test).

We then conducted a combined analysis with the 29 apes that participated in both experiments. We compared the number of first looks (maximum of two looks; i.e., one per experiment) each subject made to the target versus to the distractor during the central-approach and central-reach periods (Fig. 3). Apes made significantly more first looks to the target than to the distractor, both overall (Wilcoxon signed rank test: $Z = 3.25$, $N = 29$, $P = 0.001$, $r = 0.42$) and in each condition (FB1: $Z = 1.98$, $N = 15$, $P = 0.046$, $r = 0.36$; FB2: $Z = 2.15$, $N = 14$, $P = 0.031$, $r = 0.40$) (Fig. 3A). No significant difference was detected across species. To test this, we first calculated difference scores for each ape (number of first looks to target minus to distractor) and then subjected these scores to the Kruskal-Wallis H test [$\chi^2(2) = 0.46$, $P = 0.79$] (Fig. 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 explanations. 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, the apes 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. We confirmed that apes closely tracked all such movements (figs. S3 and 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. By requiring subjects to make predictions in situations that involved a constellation of novel features (e.g., a human attacking an apelike character hiding in a haystack), we also minimized the possibility that subjects could apply behavior rules acquired through extensive learning during past experiences. Nevertheless, 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). However, this explanatory framework cannot easily accommodate the diversity of existing evidence for ape TOM (3) nor can it account for recent evidence that human infants and apes appear to infer whether others can see through objects that look opaque, based on their own experience with the occlusive properties (i.e., see-through or opaque) of those objects (19, 20).

Thus, our results, in concert with existing data, 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 specific 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 between tasks may reflect differences in task demands or context, or less flexible abilities in apes compared with humans. At minimum, apes can anticipate that an actor will pursue a goal object where he last saw it, even though the apes 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.



Fig. 1. Events shown in experiment one.

(A to F) Familiarization. (G to J) Belief induction for the FB1 (false belief one) condition. (K) Central approach for FB1. (L to O) Belief induction for FB2. (P) Central approach for FB2. (Q) Areas of interest (AOIs) defined for the target and distractor haystacks. See movie S1.

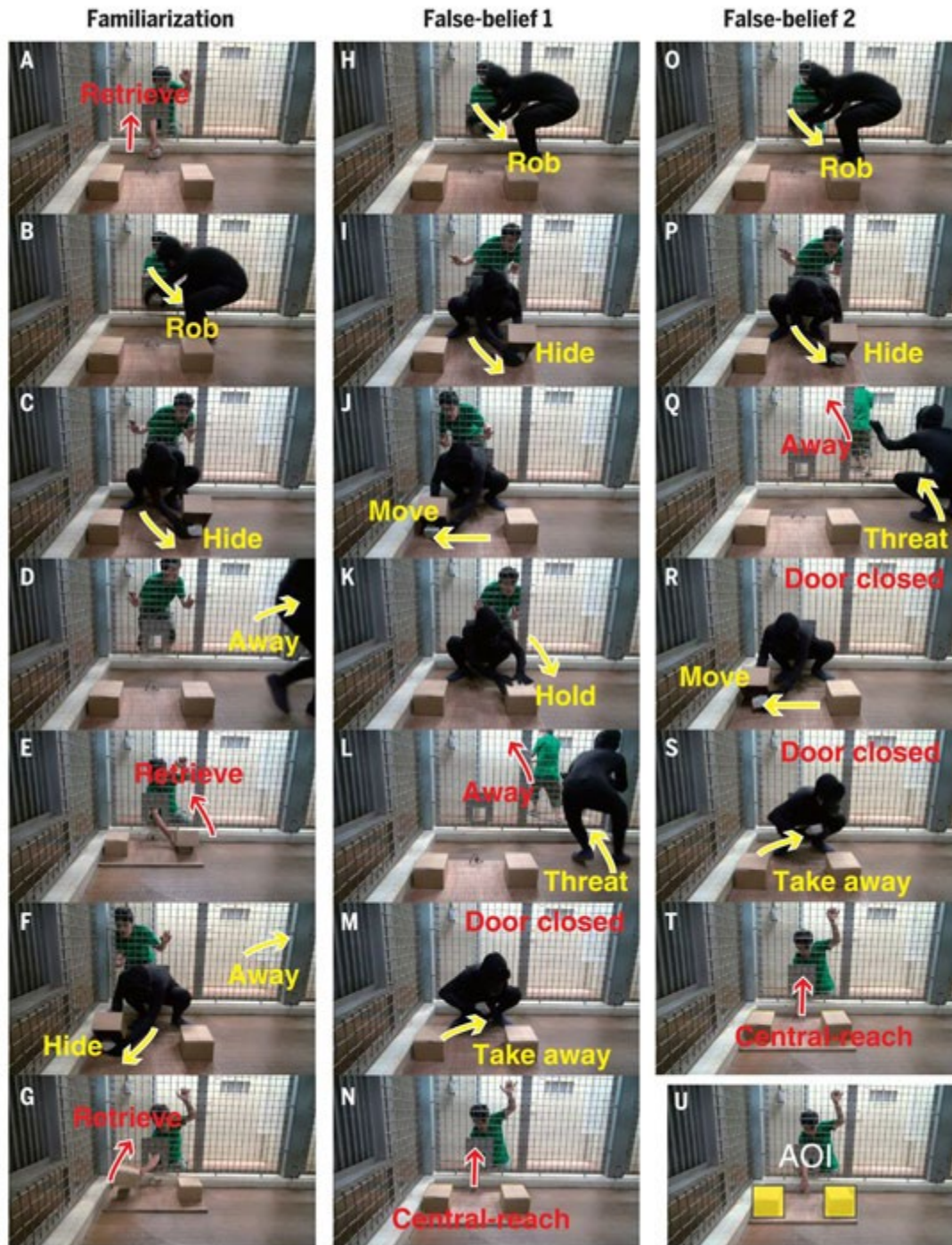


Fig. 2. Events shown in experiment two.

(A to G) Familiarization. (H to M) Belief induction for the FB1 condition. (N) Central reach for FB1. (O to S) Belief induction for FB2. (T) Central reach for FB2. (U) AOIs defined for the target and distractor boxes. 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. See movie S2.

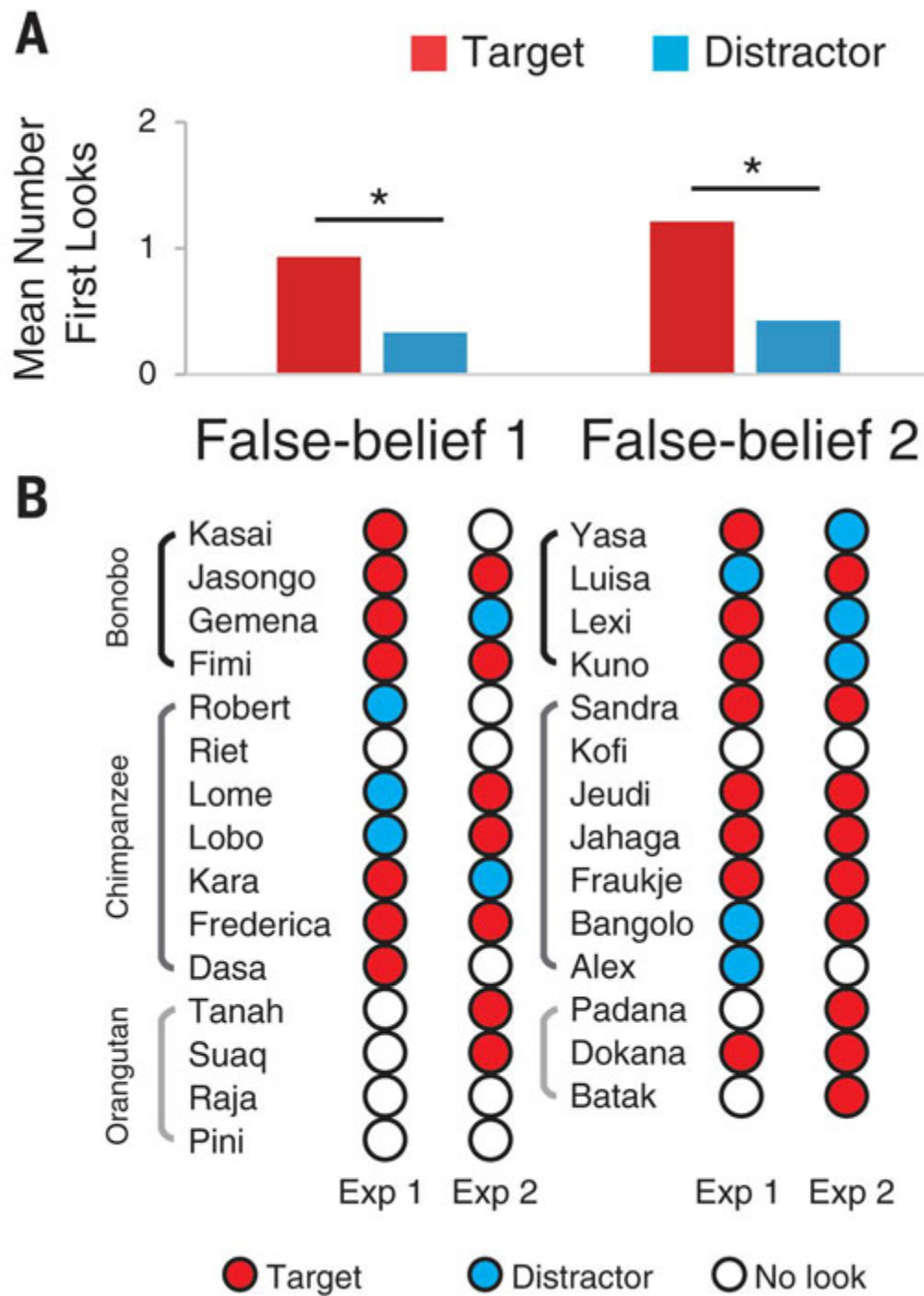


Fig. 3. Apes' performance across the two experiments. (A)

(A) Mean number of first looks to the target and the distractor for the 29 subjects who participated in both experiments. Asterisks indicate $P < 0.05$, Wilcoxon signed rank test. (B) Individual scores in each experiment.

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

Table 1. Number of participants who made first looks to either the target or the distractor during the agent's approach in experiments one ($N = 40$) and two ($N = 30$).

References and Notes

1. D. Premack, G. Woodruff, Does the chimpanzee have a theory of mind. *Behav. Brain Sci.* 1, 515–526 (1978). [doi:10.1017/S0140525X00076512](https://doi.org/10.1017/S0140525X00076512)
2. R. W. Byrne, A. W. Whiten, Machiavellian Intelligence: *Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans* (Clarendon Press, 1988).
3. J. Call, M. Tomasello, Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192 (2008). [doi:10.1016/j.tics.2008.02.010](https://doi.org/10.1016/j.tics.2008.02.010)
4. B. Hare, J. Call, M. Tomasello, Do chimpanzees know what conspecifics know? *Anim. Behav.* 61, 139–151 (2001). [doi:10.1006/anbe.2000.1518](https://doi.org/10.1006/anbe.2000.1518)
5. J. Kaminski, J. Call, M. Tomasello, Chimpanzees know what others know, but not what they believe. *Cognition* 109, 224–234 (2008). [doi:10.1016/j.cognition.2008.08.010](https://doi.org/10.1016/j.cognition.2008.08.010)
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). [doi:10.1111/j.1467-7687.2008.00793.x](https://doi.org/10.1111/j.1467-7687.2008.00793.x)
7. H. M. Wellman, D. Cross, J. Watson, Meta-analysis of theory-of-mind development: The truth about false belief. *Child Dev.* 72, 655–684 (2001). [doi:10.1111/1467-8624.00304](https://doi.org/10.1111/1467-8624.00304)
8. K. H. Onishi, R. Baillargeon, Do 15-month-old infants understand false beliefs? *Science* 308, 255–258 (2005). [doi:10.1126/science.1107621](https://doi.org/10.1126/science.1107621)
9. W. A. Clements, J. Perner, Implicit understanding of belief. *Cogn. Dev.* 9, 377–395 (1994). [doi:10.1016/0885-2014\(94\)90012-4](https://doi.org/10.1016/0885-2014(94)90012-4)
10. V. Southgate, A. Senju, G. Csibra, Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18, 587–592 (2007). [doi:10.1111/j.1467-9280.2007.01944.x](https://doi.org/10.1111/j.1467-9280.2007.01944.x)
11. D. C. Martcorena, A. M. Ruiz, C. Mukerji, A. Goddu, L. R. Santos, Monkeys represent others' knowledge but not their beliefs. *Dev. Sci.* 14, 1406–1416 (2011). [doi:10.1111/j.1467-7687.2011.01085.x](https://doi.org/10.1111/j.1467-7687.2011.01085.x)
12. A. Martin, L. R. Santos, The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition* 130, 300–308 (2014). [doi:10.1016/j.cognition.2013.11.016](https://doi.org/10.1016/j.cognition.2013.11.016)

13. F. Kano, J. Call, Great apes generate goal-based action predictions: An eye-tracking study. *Psychol. Sci.* 25, 1691–1698 (2014).[doi:10.1177/0956797614536402](https://doi.org/10.1177/0956797614536402)
14. F. Kano, S. Hirata, Great apes make anticipatory looks based on long-term memory of single events. *Curr. Biol.* 25, 2513–2517 (2015).[doi:10.1016/j.cub.2015.08.004](https://doi.org/10.1016/j.cub.2015.08.004)
15. Materials and methods are available as [supplementary materials](#)
16. J. Perner, T. Ruffman, Infants' insight into the mind: How deep? *Science* 308, 214–216 (2005).[doi:10.1126/science.1111656](https://doi.org/10.1126/science.1111656)
17. C. Heyes, False belief in infancy: A fresh look. *Dev. Sci.* 17, 647–659 (2014).[doi:10.1111/desc.12148](https://doi.org/10.1111/desc.12148)
18. R. Baillargeon, R. M. Scott, Z. He, False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118 (2010).[doi:10.1016/j.tics.2009.12.006](https://doi.org/10.1016/j.tics.2009.12.006)
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).[doi:10.1177/0956797611411584](https://doi.org/10.1177/0956797611411584)
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). [doi:10.1016/j.anbehav.2015.04.028](https://doi.org/10.1016/j.anbehav.2015.04.028)

Acknowledgments

We thank the staff at the Wolfgang Kohler Primate Research Center and Kumamoto Sanctuary for assistance. Financial support was provided by the NSF Graduate Research Fellowship Program (grant DGE-1106401 to C.K.), the Ministry of Education, Culture, Sports, Science and Technology of Japan (grant K-CONNEX to F.K.), the Japan Society for the Promotion of Science (grants KAKENHI 26885040 and 16K21108 to F.K. and grants KAKENHI 26245069 and 24000001 to S.H.), and the European Research Council (Synergy grant 609819 SOMICS to J.C.). Data are available in the main text and supplementary materials.

Supplementary Materials

Materials and Methods

Participants

A total of 41 great apes (15 bonobos, *Pan paniscus*, 19 chimpanzees, *Pan troglodytes*, 7 orangutans, *Pongo abelii*) participated in this study (Table S1). They were born in captivity and lived with conspecifics in enriched naturalistic environments at the Wolfgang Kohler Primate Research Center (WKPRC) in Leipzig, Germany, and at Kumamoto Sanctuary (KS) in Kumamoto, Japan. The apes have some experience watching naturalistic movies for enrichments and in experiments (13–14), although they were never explicitly trained for their gaze behavior.

Ethics statement

All participants were tested in the testing rooms prepared for each species, and their daily participation in this study was voluntary. They were given regular feedings, daily enrichment, and had ad libitum access to water. Animal husbandry and research protocol complied with local guidelines, which strictly adhere to international standards [the Weatherall report “The use of non-human primates in research”] and the national laws of Japan and Germany [KS: Wildlife Research Center “Guide for the Animal Research Ethics” (No. WRC-2014KS001A)] [WKPRC: “EAZA

Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria”, “WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums”, “Guidelines for the Treatment of Animals in Behavioral Research and Teaching” of the Association for the Study of Animal Behavior (ASAB)”].

Apparatus

We used the same set-up that we previously established to record apes’ eye movements accurately but non-invasively without a head-restraint device. An infrared eye-tracker was used to record their eye movements [60 Hz; X120 in WKPRC and X300 in KS; Tobii Technology AB, Stockholm, Sweden]. In each facility, we utilized the setups that were already established to test apes in this and other experiments. KS chimpanzees participated with a familiar experimenter in the same testing room. Their heads were positioned directly by the hands of the experimenter. The eye-tracker and monitor were installed outside of the testing room. Apes watched the movies on the monitor through a transparent acrylic panel (1-2 cm in thickness). We confirmed that this transparent acrylic panel does not interfere with recording of eye movements (i.e., the infrared reflection from the eyes). KS bonobos and WKPRC bonobos, chimpanzees, and orangutans participated with the experimenter on the other side of a transparent acrylic panel (i.e., not in the same testing room). We therefore installed a nozzle and a tube on the transparent acrylic panel, which kept subjects relatively stationary during testing by allowing them to sip diluted grape juice (via a custom-made juice dispenser made from a medical drip). Apes were allowed to drink juice freely while watching the movies (i.e. irrespective of their gaze behavior). In both facilities, the movies were presented at a viewing distance of 70 cm with a resolution of 1280×720 pixel on a 23-inch LCD monitor (43×24 degree) with Tobii Studio software (version 3.2.1). Their eye-movement responses (first looks and viewing times) to each scene feature (e.g. Target, Distractor) were coded automatically in the Tobii Studio software based on the Area-Of-Interest (see Data analysis for details).

Two-point automated calibration was conducted for the apes by presenting a small object or movie clip on each reference point. Each time before the recording session, we manually checked the accuracy and repeated the calibration if necessary. Calibration errors are typically within a degree.

Stimuli and procedure

In each experiment (Figure 1 and 2), each ape participant watched videos consisting of two familiarization trials and one test trial. In experiment 1, subjects experienced both familiarization trials and the test trial on the same day. In experiment 2, subjects experienced familiarization trials on one day and the test trial on the next (see below for details). The test trial presented either of two conditions: false-belief 1 (FB1) or falsebelief 2 (FB2). Subjects participated in a single trial (i.e., including two familiarization trials and one test trial) for each experiment, with half of the participants randomly assigned to FB1, and the rest to FB2 (i.e. between-subject design). Between subjects, we counterbalanced the order of familiarization trials (L then R or R then L) and, in the test trial, whether the correct location for anticipatory looks (i.e., the Target) was the left or right location. That is, we prepared four patterns of presentation for each condition that counterbalanced the sides where the object was hidden in each familiarization (L then R or R then L), as well as the side (L or R) where the Actor believed that the object was hidden in the test (i.e., four combinations: LRL, LRR, RLL, and RLR). We assigned these sequences to participants as evenly as possible.

Experiment 1 (letters refer to the panels in main text Figure 1)

On the first familiarization trial, a human actor (Actor) was positioned between two haystacks with openings in them (such that an object or agent could be hidden within either haystack). The Actor peeked inside both haystacks (to imply that the backs of the haystacks were closed, and that only from his central perspective could one see inside the haystacks) (a). The Actor then turned his back to the camera and walked toward the door in the background of the scene (b). The King-Kong

character (a person wearing a King-Kong suit; KK) entered the scene from the foreground and hit the Actor on the back. The Actor then ran through the door, grabbed a long orange stick, and came out from the door. While the Actor watched, KK ran and hid in one of the haystacks (c). The Actor raised his stick above his head, walked centrally to the middle of the two haystacks, paused, and then turned and hit with his stick the top of the haystack containing KK (i.e., Target; the other side was Distractor) (d). KK ran away and the scene faded out. On the second familiarization trial, the Actor came out from the door, and the same actions repeated except that KK hid in (e), and the Actor hit (f), the other haystack.

On the test trials for FB1, the Actor came out from the door without a stick, and watched as KK hid in one of the haystacks (Distractor) (g). The Actor continued to watch as KK moved to the other haystack (h). The Actor then ran through the door and shut it behind him (i). While the Actor was out of the scene, KK left the haystack and ran away (j). The Actor then opened the door with a stick in hand and, as in the familiarization, walked centrally to the middle of the two haystacks (k). During this central-approach, both haystacks flashed once with a ring sound to encourage the ape participants to make explicit looks to either the Target or Distractor. The Target was the haystack where the Actor had last seen KK hide and should falsely believe KK to still be (i.e., the second hiding location in FB1). The test trials for FB2 were the same as those for FB1 except that after the Actor watched KK hide in the first location (Target) (l), he exited through the door (m). Consequently, the Actor did not see when KK moved to the second location (Distractor) (n) or exited that location and ran away (o). As in FB1, after KK had left, the Actor re-emerged from the door and proceeded centrally between the haystacks (p). (total video duration = 53 sec.)

Experiment 2 (letters refer to the panels in main text Figure 2):

This experiment relied on the same design as Experiment 1 but employed a new scenario. In these scenes, the Actor was pictured in the background behind a mesh wall. During the first familiarization trial, the Actor attempted to retrieve a stone through an opening in the mesh (a). KK entered the scene on the other side of the mesh from the Actor, took the stone from the Actor (b), and agitated in front of Actor. While the Actor watched, KK then hid the stone in one of two boxes in front of Actor (c), and then framed out (d). The Actor reached ambiguously toward the two boxes (grabbed the string attached to the plate that had the two boxes on it, and dragged that plate toward himself), and then flipped over the correct box and retrieved the stone (e). KK then framed in and stole the stone. The scene faded out. During the second familiarization trial, KK had the stone, and the same actions were repeated except that KK hid the stone in (f)—and the Actor retrieve it from (g)—the box located on the other side.

During the test trial for FB1, KK hid the stone in one of the boxes (Distractor) while the Actor watched (h). KK moved the stone to the other box (Target), while the Actor was still watching (i). KK held the top of the Distractor box while the Actor watched (control for the last location that the Actor saw) (k). KK threatened the Actor, and the Actor ran through the door in the background and shut it behind him (l). KK retrieved the stone from under the box and framed-out (m). The Actor returned through the door and reached ambiguously toward the two boxes (n). During this central-reach, both boxes flashed several times with ring sounds to encourage the ape participants to make explicit looks to either the Target or Distractor. In FB2 trials, KK stole the Actor's stone (o) and hid it in one box (Target) (p) while the Actor watched. The Actor then left (q) and did not witness KK moving the stone to (r)—and removing it from (s)—the second box (Distractor) before returning to search (t).

In our pilot test for Experiment 2, we used video files that presented test trials right after familiarization trials (1 min. in duration). When we presented these video files to the apes in one facility (Kumamoto Sanctuary; 11 participants), nearly half of apes (5 out of 11) did not make

explicit looks to either the Target or Distractor, presumably because the entire duration of videos was too long to sustain their attention. Therefore, when we tested the apes at the other facility (Wolfgang Kohler Primate Research Center), we split the familiarization trials (42 sec.) and the test trial (39 sec.) across two consecutive days for each participant. The first six seconds of test trials repeated the same scenes of familiarization 1 (KK's agitation) to remind apes of the previously-shown contents. A previous study confirmed that apes can remember video content across days (14).

Data analysis

Polygon-shaped Areas-Of-Interest (AOI) were defined for the Target and Distractor haystacks and boxes (Figure 1q and Figure 2u). The AOIs of the haystacks were slightly enlarged to encompass the area just above them, where the Actor hit the top of the haystacks. Eye-movement data were filtered using a Tobii fixation filter. Our primary measure of anticipatory looking was each subject's first-look to the Target or Distractor during the central-approach (Experiment 1) and central-reach (Experiment 2) portions of the stimuli. The time-window was common across familiarization and test trials and also across Experiments 1 and 2. This window began when the Actor started walking or reaching and ended when the Actor stopped walking or reaching (4.5 sec. in duration). We conducted independent binomial tests for each experiment to determine whether subjects looked to the Target significantly above chance, and Fisher's exact test to determine whether a difference in performance existed between FB1 and FB2 conditions. These analyses followed Southgate et al. (10). In addition, we performed a combined analysis of performance across experiments for the 29 subjects who participated in both (related samples Wilcoxon signed rank test comparing the number of first looks to the Target vs to the Distractor).

Supplementary Text

Directionality of the Actor's gaze, central-approaches, and central-reaches

It is critically important that participants could not predict the Actor's actions based on the directionality of his gaze, approach, or reach. We controlled this issue in the following manners. First, during filming for both studies, the Actor looked at a marker (not visible in the final videos) that was placed directly between the Target and Distractor (Figure S1). Second, following previous ToM studies (8,10,19), the Actor wore a cap in Exp. 1, so that participants could not track any slight changes in gaze direction during the Actor's central-approach. Third, we checked to be sure that the movements of the stick or Actor's arm do not provide any directional cues (Figure S2). Finally, we asked 12 naïve human adults to watch just the central-approaches and central-reaches and to code whether they believed that the Actor was going to hit, or grab, the haystack/box on the left or right. Each coder examined all eight videos used in this study. The order of presenting videos was counterbalanced across coders. The coders couldn't identify the Target (47.9%, one-sample t-test, $t(11) = 0.56$, $p = 0.58$, in Exp. 1; 56.2%, $t(11) = 1.39$, $p = 0.19$, in Exp.2).

Familiarization results

Previous studies with human infants (10,19) only included in their false belief analyses subjects who made anticipatory looks to the Target on the second familiarization trial because young infants may vary in their tendency to make anticipatory looks. As we did not expect variation in anticipatory looking across apes based on species or age, consistent with a previous study (13), we did not restrict our analyses based on this criterion. However, importantly, following this inclusion criterion (i.e., excluding from FB analyses all subjects who did not make any first looks to the Target in familiarization 2 of both Experiment 1 and 2; $N = 7$) does not change our results. Specifically, with this restricted sample ($N = 22$), in the combined analysis of FB trials, subjects still made significantly more first looks to the Target than to the Distractor (Wilcoxon signed-rank test, Target = 0.95 looks vs. Distractor = 0.41 looks, $Z = 2.37$, $p = 0.017$, $r = 0.35$). This finding ensures that our results mirror those from studies with infants.

The purpose of the familiarization trials was to prepare subjects to anticipate the Actor's behavior by the FB test trial—by exposing them to our novel scenarios and to the Actor's tendency to correctly pursue the object in both the left and right locations. However, to avoid excessive learning or apes losing interest, we minimized the number of familiarization trials to just two. Apes tended to make first looks to the Target more than to the Distractor even on these familiarization trials. Yet, this tendency was not significant (Table S2). Presumably, some subjects required this period of familiarization before they could make accurate predictions of the Actor in our novel scenarios.

As apes in Experiment 1 also participated in a follow-up experiment involving the same stimuli (see below), we analyzed their performances on the familiarization trials of this follow-up experiment. In this follow-up experiment, we obtained similar results (Target 24 vs. Distractor 11 in familiarization 1; Target 15 vs. Distractor 16 in familiarization 2). Importantly, overall — in a combined analysis of the four familiarization trials from Experiment 1 and the follow-up replication — apes tended to make more first looks to the Target than to the Distractor (Wilcoxon signed rank test, $Z = 1.89$, $N = 40$, $p = 0.059$, $r = 0.21$, *marginally significant*). Thus, overall, apes also tended to anticipate the Actor's actions in familiarization stimuli.

Species Difference

In this study, we did not find systematic species differences in anticipatory performance. Table S3 summarizes each species' first-look performances in Experiment 1 and 2. No species consistently performed worse than the other species. In the combined analysis, we did not find a statistical difference between species (see the main Text). In the viewing-time analysis (see below; ANOVA including Species as a factor), we also did not find a significant effect of Species (except that in Experiment 1 we found a significant main effect of Species in which orangutans viewed the Target and Distractor for shorter durations than the other species). It should be noted that two previous studies that examined anticipatory looks of bonobos, chimpanzees, and orangutans also did not find any significant differences between species (13,14).

Viewing patterns during belief-induction phases

To understand the Actor's false belief and general story plots, it is essential for apes to view the Actor's and KK's actions unambiguously during the belief-induction phases. To confirm if apes indeed viewed the actions, we analyzed their viewing patterns during the belief-induction phases in Experiment 1 and 2. We first defined the Area-Of-Interest for Actor's and KK's main action areas (Figure S3A and S4A), segmented action sequences into multiple 1-sec. action scenes, and then measured the viewing times for those AOI in each action scene (Figure S3B, S3C, Figure S4B, S4C). We confirmed that Apes did view the Actor's and KK's action unambiguously in each action scene.

Viewing times in the anticipatory-look time window

Longer viewing time to Target than Distractor in the anticipatory-look time window is another measure of anticipatory looking, which has often been used in previous studies (10,13,14,19). An analysis of viewing times to Target and Distractor during the central approach and central-reach periods revealed similar results to those based on first looks. We conducted a repeated-measures ANOVA with Answer (Target/Distractor) as a within-subject factor and Condition (FB1/FB2) as a between-subject factor, for Exp. 1 and 2 and also for the combined results from Exp. 1 and 2. We also included Species (bonobo/chimpanzee/orangutan) as a between-subject factor in these analyses.

In Experiment 1 (Table S4), we did not find any significant effect of Answer or Condition, neither main nor interaction effect. We found only the main effect of Species [$F(2,23) = 5.08$, $p = 0.015$, $\eta^2 = 0.30$]; orangutans viewed the Target and Distractor for a shorter duration than the other two species. In Experiment 2 (Table S4), we found the main effect of Answer [$F(1,23) = 6.87$, $p =$

0.015, $\eta^2 = 0.23$], but not the main effect of Condition, Species, or the interaction effect. In the combined data (Figure S5), we found the main effect of Answer [$F(1,23) = 5.85$, $p = 0.024$, $\eta^2 = 0.20$], but not the main effect of Condition, Species, or the interaction effect. In both cases, subjects looked longer at the Target than at the Distractor.

Follow-up & Pilot tests

We conducted two follow-up tests for Experiment 1 and one pilot test for Experiment 2. The first follow-up test of Experiment 1 was conducted to replicate the same results using the same participants, stimuli and conditions (yet these were differently counter-balanced across participants) and strengthened the main results (see immediately below). The second follow-up test of Experiment 1 was conducted also to confirm the initial results. We used KK as a protagonist and the human actor as an antagonist. This follow-up test failed to elicit explicit looks to either Target or Distractor in a majority of participants (28 out of 41), presumably because subjects were excessively fixated on KK during the central-approaches. The pilot test of Experiment 2 also failed to elicit explicit looks in participants (5 out of 11) presumably because the overall duration of video was excessively long. This pilot test informed our decision to administer Experiment 2 familiarization trials and test trials on separate days, which successfully elicited anticipatory looks (see Stimuli and Procedure for details). The details of the first follow-up are presented below.

Experiment-1 follow-up (replication):

In Experiment 1, the binomial test of first looks during the full 4.5 second window revealed only a trend (i.e., $p < 0.1$) in which apes tended to make more first looks to the Target than Distractor. One way to unambiguously confirm this result is to replicate the same result using the same stimuli and participants. We thus tested the same participants again using the same sets of stimuli, but switched the assigned conditions for each participant (i.e. a within-subject design in combination with the initial results). In this follow-up test, we obtained similar results to those in the initial tests (Table S5). The binomial test comparing the number of participants who made first looks to the Target versus Distractor did not reach statistical significance (19 vs. 12, $p = 0.28$, both-sided). However, when the results from this follow-up test were combined with the initial data set, a Wilcoxon Signed-rank test confirmed that apes made more first looks to the Target than the Distractor during the central-approach ($Z = 2.06$, $p = 0.039$, $r = 0.23$). There was no significant difference in performance between the two FB conditions ($Z = 0.48$, $p = 0.62$, $r = 0.05$) in the combined data.

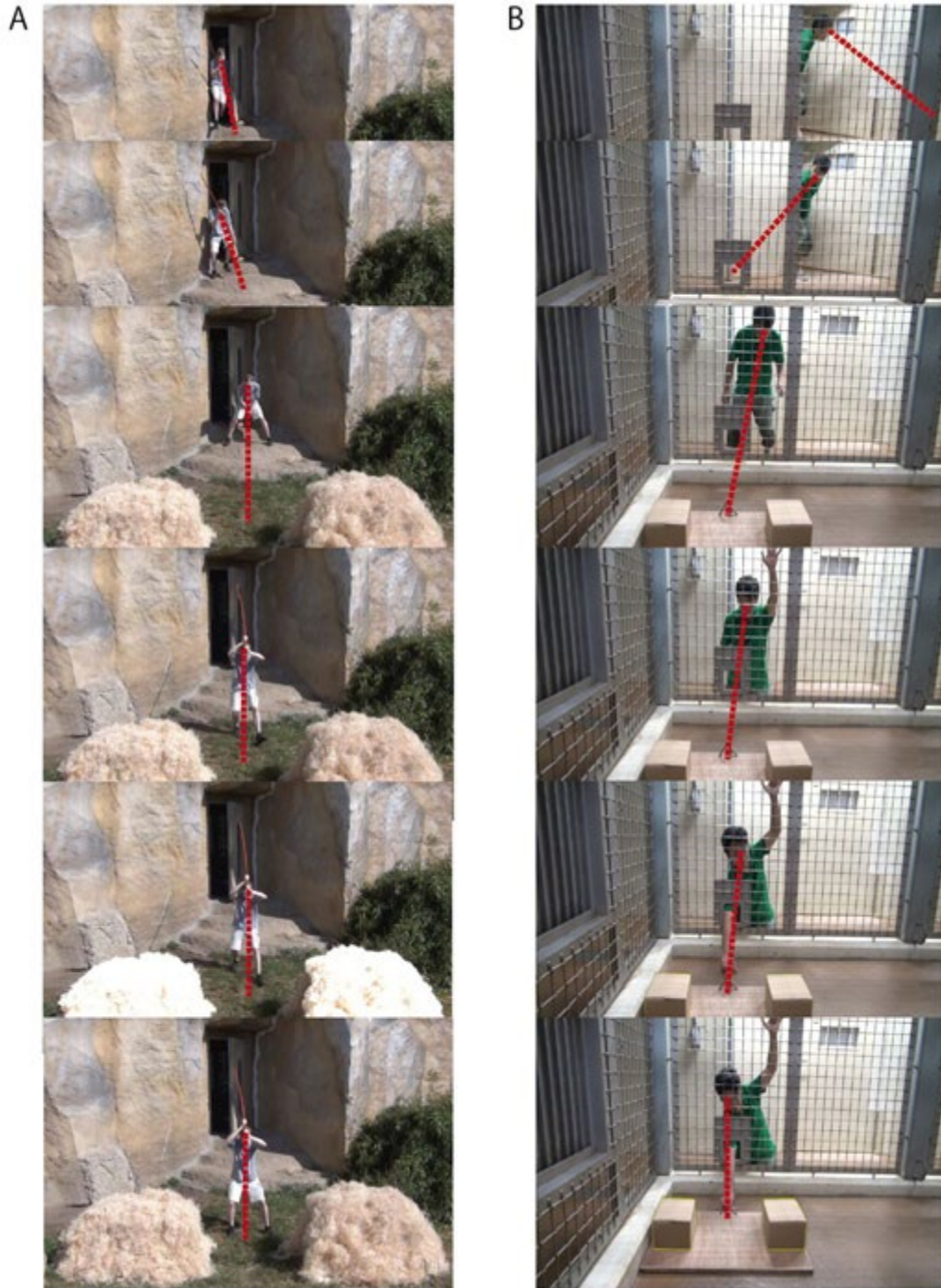


Fig. S1. Gaze control in the videos.

A. In Experiment 1, when the Actor came out from the door, he looked at the floor nearby. The Actor then shifted his gaze directly to a marker placed in between the Target and Distractor. Also note that the Actor wore a cap to hide his eye region. B. In Experiment 2, when the Actor came out from the door, he first looked at the direction where KK ran away and then looked at the hole on the mesh. He then shifted his gaze directly to a marker placed in between the Target and Distractor.

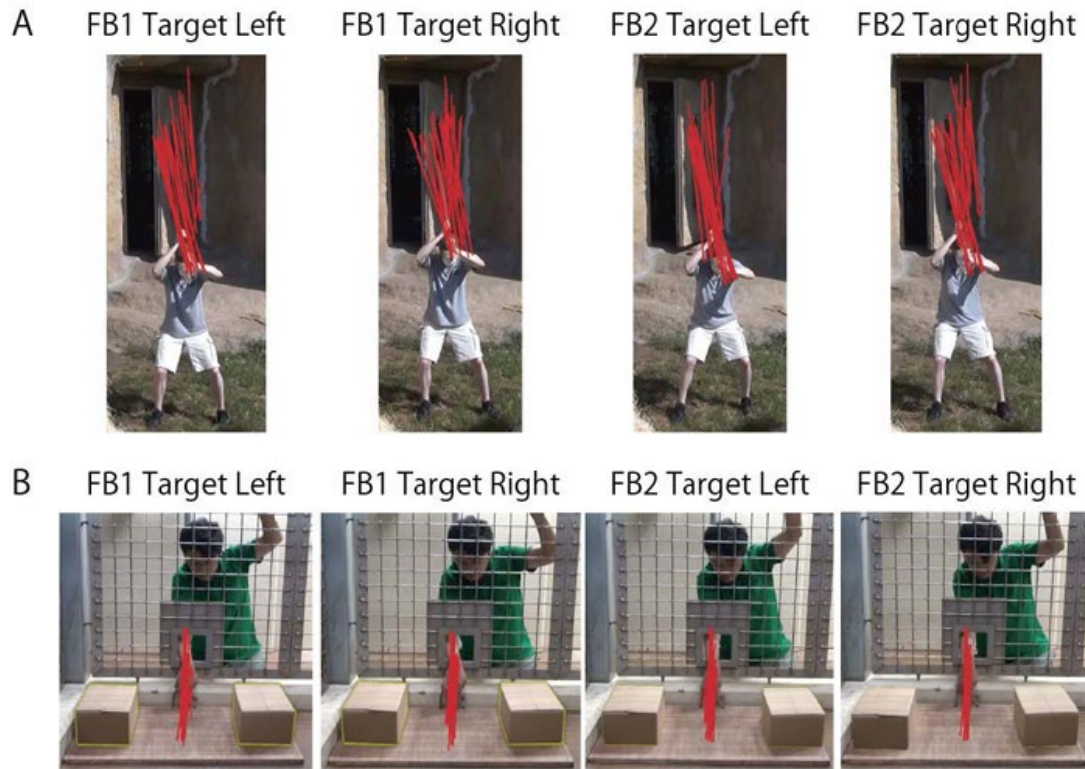


Fig. S2. Action control in the videos.

Red lines indicate the positions of the stick (Experiment 1; A) and of the Actor's arm (Experiment 2; B) during the central-approach and the central-reach, respectively (drawn 5 times per second during the 4.5-sec. time window for each of the eight videos used in this study).

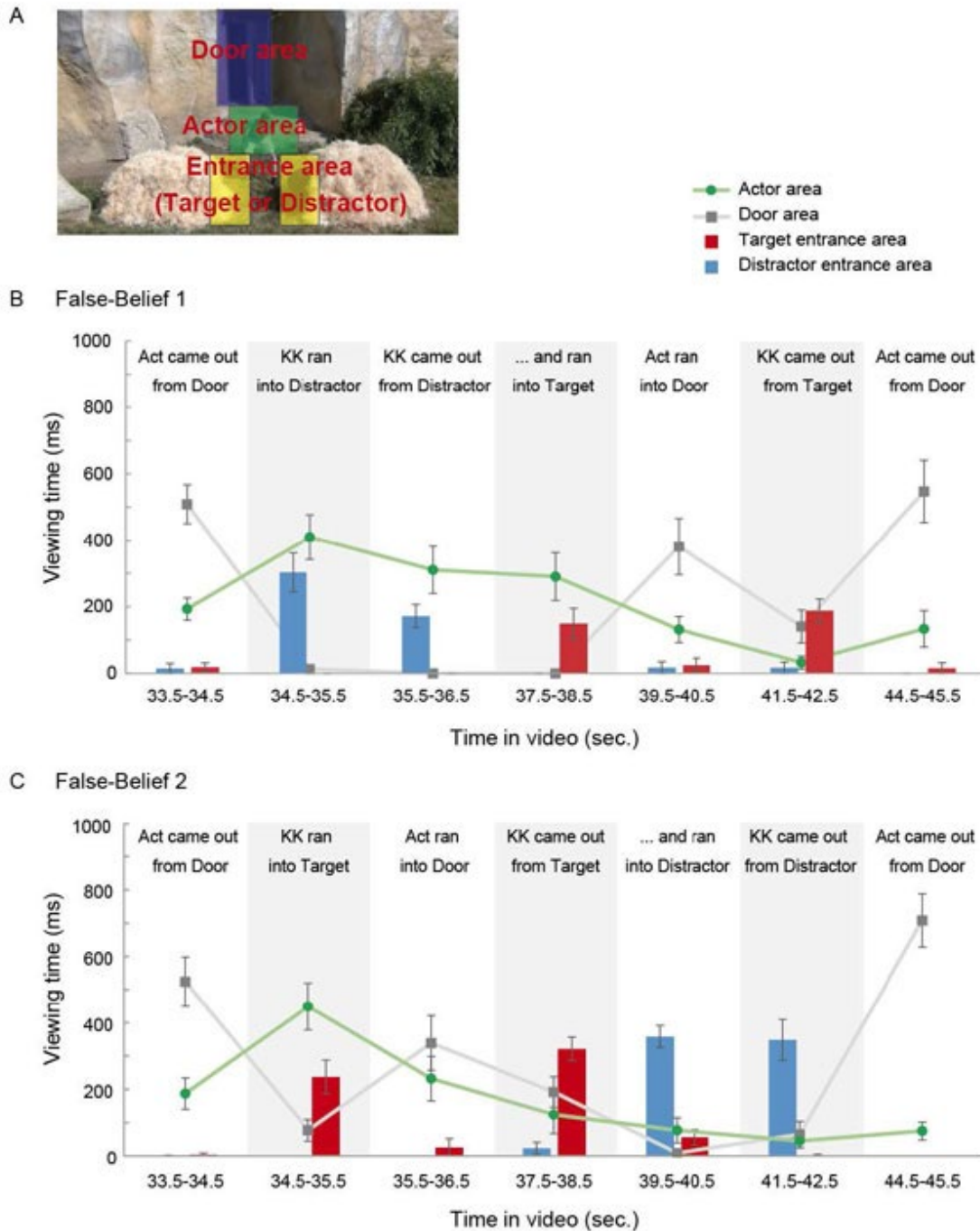


Fig. S3. Viewing patterns during the belief-induction phase in Experiment 1.

Viewing times for each Area-Of-Interest (Actor area, Door area, Target/Distractor entrance area) in each action scene (defined as a 1-sec time-window).

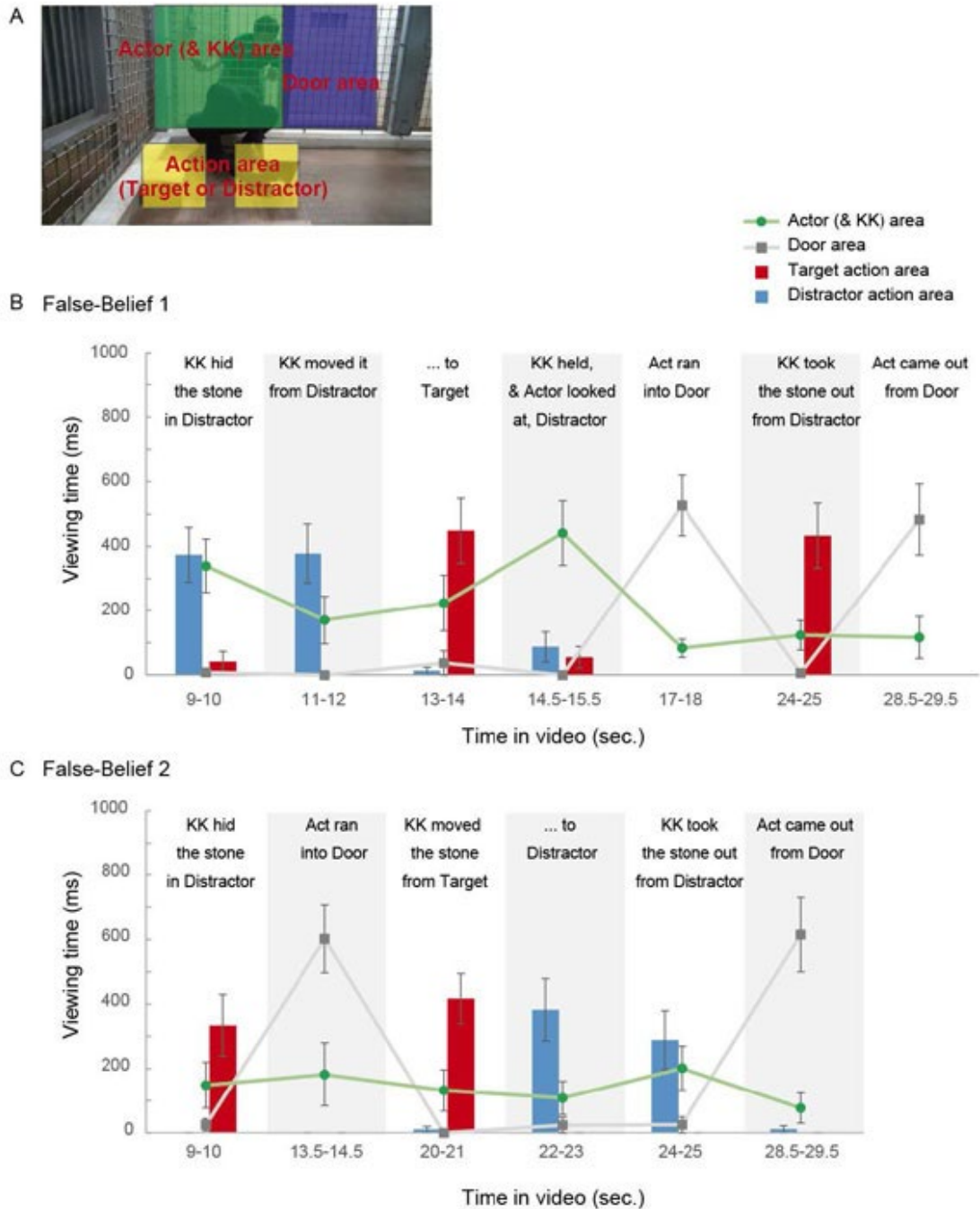


Fig. S4. Viewing patterns during the belief-induction phase in Experiment 2.

Viewing times for each Area-Of-Interest (Actor area, Door area, Target/Distractor action area) in each action scene (defined as a 1-sec time-window).

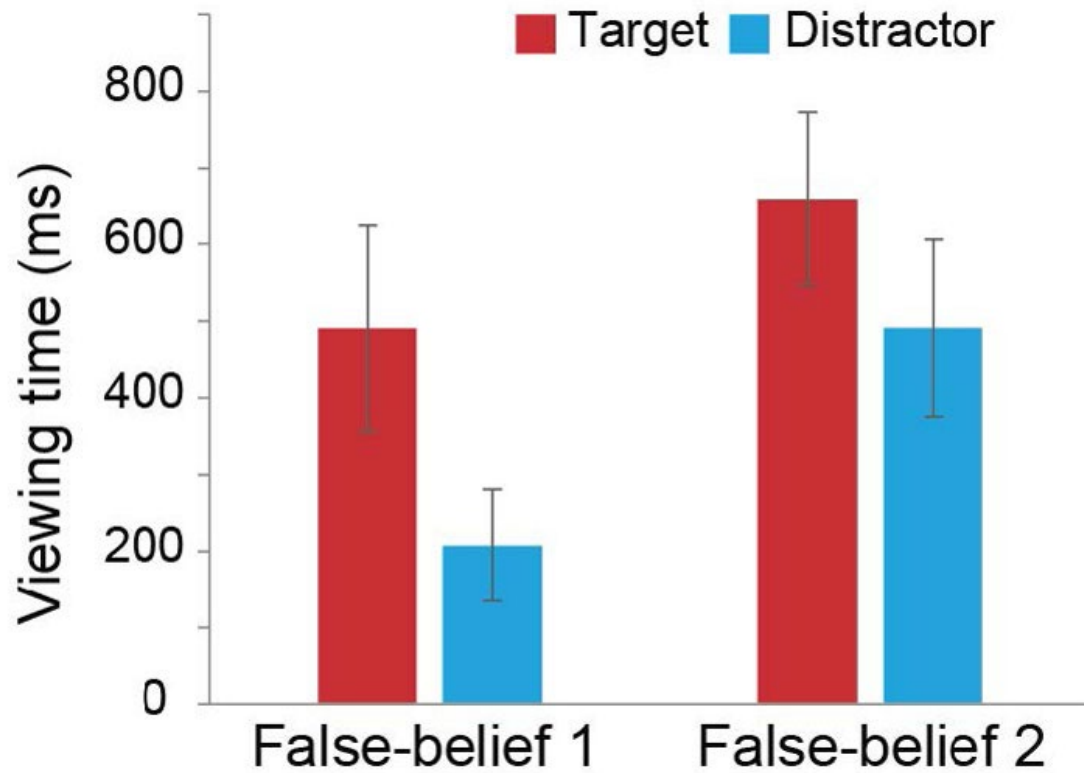


Fig. S5. The combined data for viewing-time results.

The means of total viewing times (ms \pm SE) to the Target and Distractor during the central-approach (Exp. 1) and central-reach (Exp. 2).

Subject	Facility	Species	Sex	Age class	Rearing history	Exp. 1	Exp. 2
Connie-Lenoire	KS	Bonobo	Female	Adult	Mother	FB 1	Pilot
Ikela	KS	Bonobo	Female	Adult	Nursery	FB 1	Pilot
Junior	KS	Bonobo	Male	Adult	Mother	FB 1	Pilot
Lolita	KS	Bonobo	Female	Adult	Nursery	FB 2	Pilot
Louise	KS	Bonobo	Female	Adult	Nursery	FB 2	Pilot
Vijay	KS	Bonobo	Male	Adult	Nursery	FB 2	Pilot
Hatsuka	KS	Chimpanzee	Female	Juvenile	Nursery	FB 1	Pilot
Iroha	KS	Chimpanzee	Female	Adult	Mother	FB 1	Pilot
Misaki	KS	Chimpanzee	Female	Adult	Mother	FB 2	Pilot
Mizuki	KS	Chimpanzee	Female	Adult	Nursery	FB 2	Pilot
Natsuki	KS	Chimpanzee	Female	Adult	Mother	FB 2	Pilot
Fimi	WKPRC	Bonobo	Female	Juvenile	Mother	FB 1	FB1
Gemena	WKPRC	Bonobo	Female	Adult	Mother	FB 1	FB1
Jasongo	WKPRC	Bonobo	Male	Adult	Mother	FB 1	FB1

Subject	Facility	Species	Sex	Age class	Rearing history	Exp. 1	Exp. 2
Kasai	WKPRC	Bonobo	Male	Juvenile	Mother	FB 1	FB1
Kuno	WKPRC	Bonobo	Male	Adult	Nursery	FB 2	FB2
Lexi	WKPRC	Bonobo	Female	Adult	Nursery	FB 2	FB2
Luisa	WKPRC	Bonobo	Female	Adult	Mother	FB 2	FB2
Yaro	WKPRC	Bonobo	Male	Juvenile	Mother	*	FB1
Yasa	WKPRC	Bonobo	Female	Adult	Mother	FB 2	FB2
Alex	WKPRC	Chimpanzee	Male	Adult	Nursery	FB 2	FB2
Bangolo	WKPRC	Chimpanzee	Male	Juvenile	Mother	FB 2	FB2
Dasa	WKPRC	Chimpanzee	Female	Adult	Nursery	FB 1	FB1
Fraukje	WKPRC	Chimpanzee	Female	Adult	Nursery	FB 2	FB2
Frederica	WKPRC	Chimpanzee	Female	Adult	Nursery	FB 1	FB1
Jahaga	WKPRC	Chimpanzee	Female	Adult	Mother	FB 2	FB2
Jeudi	WKPRC	Chimpanzee	Female	Adult	Mother	FB 2	FB2
Kara	WKPRC	Chimpanzee	Female	Adult	Mother	FB 1	FB1
Kofi	WKPRC	Chimpanzee	Male	Adult	Mother	FB 2	FB2
Lobo	WKPRC	Chimpanzee	Male	Adult	Mother	FB 1	FB1
Lome	WKPRC	Chimpanzee	Male	Adult	Mother	FB 1	FB1
Riet	WKPRC	Chimpanzee	Female	Adult	Nursery	FB 1	FB1
Robert	WKPRC	Chimpanzee	Male	Adult	Nursery	FB 1	FB1
Sandra	WKPRC	Chimpanzee	Female	Adult	Mother	FB 2	FB2
Batak	WKPRC	Orangutan	Male	Juvenile	Mother	FB 2	FB2
Dokana	WKPRC	Orangutan	Female	Adult	Mother	FB 2	FB2
Padana	WKPRC	Orangutan	Female	Adult	Mother	FB 2	FB2
Pini	WKPRC	Orangutan	Female	Adult	Nursery	FB 1	FB1
Raja	WKPRC	Orangutan	Female	Adult	Mother	FB 1	FB1
Suaq	WKPRC	Orangutan	Male	Juvenile	Mother	FB 1	FB1
Tanah	WKPRC	Orangutan	Female	Juvenile	Mother	FB 1	FB1

Table S1. Participant information.

Also shown are subjects' conditions in Exp. 1 and 2.

*not available when Experiment 1 was conducted

		Target	Distractor	Total
Experiment 1	Familiarization 1	21	14	35 (5)
	Familiarization 2	18	15	32 (8)
Experiment 2	Familiarization 1	13	13	26 (4)
	Familiarization 2	16	10	26 (4)

Table S2. Familiarization results.

Number of participants who made first looks to the Target and Distractor during the central-approach (Experiment 1) or central-reach (Experiment 2) in the 1st and 2nd familiarizations. Shown in parentheses is the number of participants who did not look to either.

			Target	Distractor	Total
Experiment 1	FB1	Bonobo	5	1	6 (1)
		Chimpanzee	5	3	8 (1)
		Orangutan	0	0	0 (4)
	FB2	Bonobo	4	2	6 (1)
		Chimpanzee	5	4	9 (1)
		Orangutan	1	0	1 (2)
Experiment 2	FB1	Bonobo	3	1	4 (1)
		Chimpanzee	3	1	4 (3)
		Orangutan	2	0	2 (2)
	FB2	Bonobo	1	3	4 (0)
		Chimpanzee	5	0	5 (2)
		Orangutan	3	0	3 (0)

Table S3. First-look results by species

Number of participants (by species) who made first looks to either the Target or Distractor during the agent's approach in experiments 1 and 2. Shown in parentheses is the number of participants who did not look at either.

		Target	Distractor
Experiment 1	FB1	270 (86)	148 (61)
	FB2	339 (95)	335 (85)
Experiment 2	FB1	271 (136)	136 (57)
	FB2	341 (60)	107 (35)

Table S4. Viewing times in the anticipatory-look time window.

The mean viewing times (ms) to the Target and Distractor during the central-approach (Exp. 1) and central-reach (Exp. 2). Shown in parentheses is the standard error of the mean.

	Target	Distractor	Total
Familiarization 1	24	11	35 (5)
Familiarization 2	15	16	31 (9)
FB1	11	5	16 (4)
FB2	8	7	15 (5)
Total (FB1 + FB2)	19	12	31 (9)

Table S5. Results from the Experiment 1 Replication.

Number of participants who made first looks to either the Target or Distractor during the central-approach. Shown in parenthesis is the number of participants who did not look to either.