Great apes generate goal-based action predictions by eye movements

Fumihiro Kano (1, 2)
Josep Call (1, 3)

1. Department of Developmental and Comparative Psychology, Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany
2. Japan Society for Promotion of Science, Tokyo, Japan
3. School of Psychology and Neuroscience, University of St Andrews, UK

Corresponding Author’s address
fkano@gmail.com
Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Abstract

To examine great apes' online prediction of other's actions, this study used an eye-tracking technique and an experimental paradigm previously used to test human infants. Twenty-two great apes including bonobos, chimpanzees, and orangutans were familiarized to movie clips of a human hand reaching to grasp one of two objects. Then object locations were swapped, and the hand made an incomplete reach between the objects. In a control condition, a mechanical claw performed the same action. Apes predictively looked at the familiarized goal object rather than the familiarized location when viewing the hand action. However, they did make no prediction when viewing the claw action. These results were similar to those reported previously in human infants, and there was no species difference among great apes. Thus, great apes make online goal-based prediction about the other's action, which is not a unique skill among humans but is shared more widely among primates.

Key words: action prediction; eye-tracking; nonhuman primates; proactive goal-directed eye movements
Introduction

Our eyes move proactively rather than reactively when we perform manual action as well as when we observe manual action performed by the other individual (Falck-Ytter, Gredebäck, & Von Hofsten, 2006; Flanagan & Johansson, 2003; Johansson, Westling, Bäckström, & Flanagan, 2001; Land & Furneaux, 1997). That is, our eyes spontaneously seek for the goal object which is slightly ahead of our and the other’s action. This online spontaneous prediction about the other’s action should have immense functions in our dynamic social world. For example, it helps us to perform smooth coordination with others and also to have an advantage in competition with others. Also, by similarly attending to the other’s and own action, we may encode the other’s action in a similar way with our own action and thereby enhance the understandings of the other’s action.

Ontogenetically, this proactive goal-directed eye movement during action observation emerges around the 6 month of age in humans (Falck-Ytter, et al., 2006; Kanakogi & Itakura, 2011). Importantly, by the first year of life, those online predictions are performed in a non-reflexive, goal-based manner. For example, in Cannon and Woodward (2012), when 11-month-old infants had seen a person reach for two objects, they looked predictively toward the object for which the person had previously reached, even after the objects’ locations had changed (an adaptation of Woodward (1998) paradigm for an eye-tracking task). In Southgate, Senju, & Csibra (2007), when 2-year-old children had seen a person approach two containers, they looked predictively toward the one in which the person had previously seen a toy hidden, even though the toy was no longer there. Therefore, during the course of development, humans acquire a sophisticated skill in which they visually predict the other’s actions based on the underlying goals and intentions.

Evidence of such proactive goal-directed eye movements in nonhuman species is
 relatively sparse (Kano & Tomonaga, 2013; Myowa-Yamakoshi, Scola, & Hirata, 2012).
Kano and Tomonaga (2013) examined the eye movements of a chimpanzee while she was
performing a manual imitation task with an experimenter in a real life setting. They found
that the chimpanzee predictively looked at the reward when an experimenter was reaching
for it, while she reactively looked at the goal of reaching when the experimenter was
performing the task gestures. Myowa-Yamakoshi et al. (2012) found that chimpanzees,
12-month-old human infants and adults predictively looked at the goals of various actions
presented in the movies, while 8-month-old human infants reactively looked at them. Thus,
chimpanzees seem to predict the goals of some actions spontaneously as humans do.
However, it remains unclear whether such eye movements are goal-based; i.e. not simply
oriented toward the direction of hand/object movements.

There are numerous behavioral studies showing that nonhuman primates make
predictions about the other’s actions based on the perception and knowledge. For example,
when a subordinate and a dominant chimpanzee were competing for the two foods, the
subordinate avoided the food which the dominant can see and instead chose the food which
the dominant cannot see (Hare, Call, & Tomasello, 2000). In addition, chimpanzees helped
a human or conspecific partner by taking the object which the partner was trying to reach
(Warneken & Tomasello, 2006), possibly by identifying the partner’s current needs
(Yamamoto, Humle, & Tanaka, 2012). In a study using a looking time measure, when
macaque monkeys were habituated to a human experimenter reaching for an object behind
a barrier, they were more surprised to see the same yet inefficient movements of arms than
vice versa after the barrier was removed (Rochat, Serra, Fadiga, & Gallese, 2008); also see
Uller (2004). For most of the studies mentioned above, similar results were obtained in
human infants. Therefore, like humans, nonhuman primates seem to understand the other’s
actions not just in terms of surface behaviors but also in terms of the underlying goals and intentions (Call & Tomasello, 2008).

However, an important outstanding question is whether nonhuman primates make online goal-based prediction by eye movements when observing the other’s action. The aim of the current study was to investigate this aspect in three great ape species (chimpanzees, bonobos, and orangutans) with different degrees of phylogenetic distance to humans. All great ape species show basic similarities in eye movements and scene scanning (Kano, Call, & Tomonaga, 2012; Kano, Hirata, Call, & Tomonaga, 2011). Since previous studies have reported evidence of goal attribution in chimpanzees, we anticipated that chimpanzees would produce goal-based action prediction. Much less is known about the other species, with one study reporting positive results for orangutans, but not for bonobos (Buttelmann, Carpenter, Call, & Tomasello, 2007, 2008). Therefore, it is unclear whether these other species would spontaneously produce goal-based prediction when observing the other’s action, which seems to be a basic skill for general goal attribution.

We used the eye-tracking technique and adopted an experimental paradigm previously used with human infants (Cannon & Woodward, 2012). Apes were familiarized to a repeated reaching action directed to one of two objects by a human agent. The locations of the objects were then swapped, and apes’ predictive looks to the prior location versus the prior goal were examined as the agent made an incomplete reach between the objects (Fig. 1). If apes make goal-based predictions, then we expected that they would look predictively toward the prior goal, rather than to the prior location, like human infants. We also included a control condition in which a mechanical claw moved in a similar manner to the hand. The previous study with infants (Cannon & Woodward, 2012) found that infants did not predict the goal of claw actions during the test event, although the grasping action of both claw and
hand caught their attention strongly during familiarization. Thus, this control allowed us to rule out the possibility that participants looked at the prior goal during the test event simply because they attended to a salient grasping action and thereby formed an association between the hand/claw and object during familiarization.
Figure 1. Procedure of a trial. After the three familiarization events in which the hand or claw reaches for and grasps the target, the locations of objects were swapped, and then the hand or claw reaches straight between the two objects and pauses in front of, and equidistantly from the two objects.
Methods

Participants. Twenty-two great apes participated in this study (4 bonobos, *Pan paniscus*, 12 chimpanzees, *Pan troglodytes*, 6 orangutans, *Pongo abelii*; 12 females, 10 males; mean 16.0 ± S.D. 11.5 years of age; for the details, see Table S1). Five additional apes were tested but excluded from the analysis because they attended insufficiently to the stimuli (excessive off-screen fixations (n = 4); excessive eccentric eye movements (n = 1)). All apes live with their conspecifics in semi-natural indoor and outdoor enclosures at the Wolfgang Köhler Primate Research Center (WKPRC). All apes were tested in testing rooms at the WKPRC, and their daily participations of experiments were voluntary. They were given regular feedings, daily enrichment and water ad lib. Animal husbandry and research complied with the EAZA Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquariums and the WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums, respectively.

Apparatus. The eye movements of apes were non-invasively recorded with an infrared eye-tracker (60 Hz; Tobii X120, Tobii Technology AB, Stockholm, Sweden). Stimuli were presented using Tobii Studio software in a 22-inch LCD monitor (1366×768 pixel) at a 60-cm viewing distance (1 degree of gaze angle corresponded to approximately 1 cm on the monitor). We tested them unrestrained but separated from the experimenter and eye-tracker with a transparent acrylic panel. However, in order to keep their heads relatively still, we implemented a nozzle and tube attached to the acrylic panels, which produce a dripping of grape juice, and let the apes suck the nozzle during the recording (see Fig. S1). No explicit training was conducted for the apes.
Calibration. Two-point automated calibration was conducted by presenting a small object on each reference point. Relatively small numbers of reference points were used in this study because the apes tended to view those reference points only shortly. However, we manually checked the accuracy at five points after the initial calibration and repeated the calibration if necessary. As a result, we obtained the comparable accuracy between ape and human participants; a validation session with more than 10 apes and 10 humans under standard calibration results confirmed that the positional error values estimated at the five spots on the screen were less than 1-2 degree. Before every test session, we checked the accuracy manually and started the session when we confirmed the error value less than 1-2 degree around center on the screen.

Stimuli and procedure. We created our videos by slightly modifying those used in the previous study with infants (Cannon & Woodward, 2012). Overall, we slightly shortened the length of video (speeding up the movements) to match the apes’ relatively rapid shift of attention (Kano, et al., 2011). On each trial, we presented a 13160 millisecond (ms) video (resolution 1280×720 pixel) on the center of monitor, which contained a rubber toy duck (yellow) and frog (green) in the scene (Fig. 1). The video comprised three familiarization events (2230 ms each), one swap event (2550 ms), and one test event (2920 ms). During the familiarization event, either a human left hand (hand condition) or a plastic rod with claw (claw condition) appeared in the mid-right corner of scene (0-680 ms), reached to (reaching phase; 681-1210 ms), and grasped (grasping phase; 1211-2230 ms) one of the objects (target). The hand or claw moved by following a curvilinear path from the start to the object. Grasping the frog and duck accompanied a croak and quack sound, respectively. During the swap event, the other two hands (left and right hands) grasped the two objects
simultaneously and swapped the locations of them. Finally, during the test event, either a hand or a claw appeared (0-1020 ms), reached straight, and paused in front of, and equidistantly from the two objects (1021-2919 ms).

Each ape viewed one video (one trial) per day, for a total of 4 videos of hand condition and 4 videos of claw condition (i.e. within-subject design) (8 trials, over 8 days). Each ape competed either hand or claw condition first and then completed the other condition, and its order was counterbalanced across apes. Also, the object location (up or down) or the object type (duck or frog) during the familiarization was counterbalanced across apes. To prevent the hand orientation (the thumb downward/upward) from serving as a cue to predict the hand action during the test event, we used the same hand during the familiarization and test events. Therefore, if the participants learned from the familiarization event to respond to thumb orientation, their action prediction during the test events should be based on the location but not on the object (after swapping the locations of objects). The initial hand orientation in relation to the target was counterbalanced across apes.

The experimenter initiated the presentation of each video when apes were attending to the monitor. In those cases where apes moved away from the monitor during the recording (thereby severely disrupted the eye-tracking signals), we repeated the same video on the next day. This occurred only once in 4 apes. All apes completed all trials.

**Data analysis.** For the familiarization events, the apes’ predictive look was estimated by (1) the viewing time during the reaching phase (from the onset of hand/claw movement to the touch to the target). We also examined (2) the viewing time for each object during the grasping phase (from the touch to the target to the end of each familiarization event) in order to examine whether the participants similarly attended to the target and distractor (i.e. the
other object) in both hand and claw conditions. For the test events, the apes’ predictive look was estimated by (1) the viewing time during reaching (from the onset of hand/claw movements to the end of video). As there was a slight variation of phase durations among stimulus videos (familiarization reaching phase; 506 ms ± 110; familiarization grasping phase; 1023 ms ± 107 test reaching phase; 1899 ms ± 51, mean duration ± SD), the viewing times were rescaled to the values proportional to the mean duration of each phase. Apes viewed off-screen areas for 22 % ± 10.3 (mean ± SD) of total video duration. When the ape fixated off-screen areas for the entire duration of each phase, we coded those fixations as missing values (4.5 % of all data). The apes’ predictive look was additionally estimated by (2) the proportion of trials in which apes looked at the target first (vs. the distractor first) during the test event. We analyzed only the trials in which apes viewed either target or distractor (55 % of trials). Two apes and one ape did not view the objects in the hand and claw condition across 4 trials, respectively, and we excluded them from the analysis (i.e. we analyzed 19 apes for this measure).

A square-shaped AOI, 200×200 pixels in size, was defined for each of the target and distractor (both approximately 150×150 pixels in size). In addition, an AOI was defined for the trajectory area in which the hand or claw moved (Fig. S2). The fixation filtering was conducted using the Tobii fixation filter (version 3.2.1). The above-mentioned measurements were calculated using the Tobii Studio and Matlab (MathWorks, USA). Statistical analyses were conducted in SPSS version 20.
Results

Familiarization. We conducted our initial analysis for the viewing time, using a repeated-measures ANOVA with Condition (hand, claw), Object (target, distractor) as within-subject factors and Species (bonobo, chimpanzee, orangutan) as between-subject factors, respectively for reaching and grasping phase during the familiarization (averaged over three events; Fig. 2a). We confirmed that the object location (up or down) or the object type (duck or frog) which was included into the ANOVA did not significantly influence the apes’ viewing time during the familiarization, neither the main effect nor interactions with the other factors, neither during the reaching nor grasping phase (ps > 0.23). We therefore did not include this counterbalanced factor in the subsequent analyses.

We found a significant interaction between Condition and Object in the viewing time when the hand or claw was reaching for the target during the familiarization \[F(1,19) = 8.42, p = 0.009, \eta^2 = 0.30\]. We also found the main effect of Species \[F(2,19) = 3.87, p = 0.039, \eta^2 = 0.29\]. The other main effects or interactions were not significant. Post-hoc tests revealed that apes viewed the target for a longer time than the distractor in hand condition \[t(21) = 2.58, p = 0.017, \text{Cohen’s } d = 0.57\] but not in the claw condition \[t(21) = 0.51, p = 0.61, \text{Cohen’s } d = 0.16\] during this period. Also, there was a significant difference between conditions in the viewing time for the target \[t(21) = 2.64, p = 0.015, \text{Cohen’s } d = 0.56\] but not for the distractor \[t(21) = 0.03, p = 0.97, \text{Cohen’s } d = 0.007\]. Thus, these results revealed that apes predicted the target of a hand action but not that of a claw action during the familiarization.

To examine the presence of predictive saccades to the target, we measured the timing of gaze arrival at the target relative to the start of grasping (the end of reach) in the first of three familiarization events. On average, apes viewed the target 12 ms (SD = 411)
before and 213 ms (SD = 771) after the start of first grasping, in the hand and claw condition, respectively. Compared with the ordinary saccadic reaction time in these species (Kano, et al., 2011), minimum 200 ms, they viewed the hand action proactively [t(21) = 2.41, p = 0.025, Cohen’s d = 0.51], while they did the claw action reactively [t(21) = 0.08, p = 0.93, Cohen’s d = 0.01].

In contrast to the reaching phase, there was no interaction between Condition and Goal in the viewing time when the hand or claw was grasping the target during the familiarization [F(1,19) = 1.17, p = 0.29, η² = 0.058]. Thus, apes attended to the target versus distractor with the claw and hand similarly during this period. We also found the interaction between Object and Species [F(2, 19) = 5.39, p = 0.014, η² = 0.36] and the main effects of Object [F(1,19) = 143.01, p < 0.001, η² = 0.88] and Species [F(2, 19) = 6.31, p = 0.008, η² = 0.39]. The other main effects or interactions were not significant.

Test. Crucially, during the test event, we found a significant interaction between Condition and Object in the viewing time when the hand or claw was reaching between the two objects [F(1,19) = 7.48, p = 0.013, η² = 0.28]. We also found the main effect of Species [F(2,19) = 7.93, p = 0.003, η² = 0.45]. The other main effects or interactions were not significant.

Post-hoc tests revealed that apes viewed the target for a longer time than the distractor in the hand condition [t(21) = 2.50, p = 0.020, Cohen’s d = 0.53] but not in the claw condition [t(21) = 1.46, p = 0.15, Cohen’s d = 0.27] during this period. Also, there was a significant difference between conditions in the viewing time for the target [t(21) = 3.75, p = 0.001, Cohen’s d = 0.80] but not for the distractor [t(21) = 0.68, p = 0.50, Cohen’s d = 0.14]. The order of presentation (hand or claw) had no significant effect when we included it into the ANOVA (neither a main effect nor interactions with the other factors; ps > 0.22). Thus, these
results show that apes predicted the target of hand action based on the object but not on the location.

Including the factor Trial (1st-4th) in the analysis did not change the results. In addition, although apes viewed the claw trajectory generally longer than the hand trajectory, their viewing times for the objects (i.e. target plus distractor) were similar between the claw and hand conditions (Fig. S3).

To examine the initial responses to the target/distractor when the hand or claw starts reaching, we examined the proportion of trials in which apes looked at the target first (vs. the distractor first) (Fig. 2b). We found that apes’ first look to the target was more frequent in the hand than claw condition \( t(18) = 2.19, p = 0.041 \). Post-hoc tests revealed that the first look to the target was more frequent than chance (0.5) in the hand condition \( t(20) = 3.47, p = 0.002 \) but not in the claw condition \( t(19) = 0.15, p = 0.88 \).

Species difference. For the analyses mentioned above, species differences emerged in the strength of viewing the objects (i.e. the main effect of Species in the above analyses) but not in the pattern of predictive look (i.e. the interaction between Species and Condition/Object).

In general, orangutans looked at the objects longer (thus looked at the trajectory area shorter) than the other two species (see Fig. S4). Including sex (male, female) and age (young ≤ 9, adult > 9 years of age) into the analyses did not change the results.
Figure 2. (a) Viewing time (ms) when the hand or claw was reaching for the target (duration: 506 ms) and grasping the target (dur. 1023 ms) during the familiarization or was reaching between the two objects (dur. 1899 ms) during the test events. (b) Proportion of trials in which the apes looked at the target first (vs. the distractor first) during the test events. Error bars denote 95% confidence intervals. * p < 0.05, ** p < 0.01, *** p < 0.001.
Great apes’ eye movements were proactive when viewing the reaches of a person but reactive when viewing the reaches of a mechanical claw. Such proactive eye movements were goal-directed, and not simply oriented toward the direction of movements. That is, after apes viewed a person reaching to grasp one of the two target (familiarization event), and then saw that the objects’ locations were swapped (swap event), they predicted that the person’s subsequent reaches would be directed to the prior goal (test event). In contrast to this hand condition, they did not make any prediction when viewing the same actions of claw during the test event.

Great apes’ goal-based prediction about hand action is unlikely to result from simple learning of the entire action sequence because apes never viewed completed hand actions once the target had been moved during the test event. It is also unlikely to result from a simple association of grasping action and object during the familiarization because they similarly and strongly attended to the target being grasped in both hand and claw conditions. Finally, the absence of goal-based prediction about claw action is unlikely to result from their excessive attention to the unfamiliar claw stimulus (i.e. the failure to disengage attention from the claw itself) because apes viewed both target and distractor in both hand and claw conditions equally and, furthermore differentially viewed the target versus distractor in the former but not in the latter condition (i.e. Object × Condition). Thus, apes’ goal-based action prediction seemed to depend on the familiarity of the agents’ goal-directed behaviors rather than the agents’ saliency or movement per se.

The current findings are strikingly similar to those of human infants reported previously (Cannon & Woodward, 2012). Also, in this study, although ape species (bonobos, chimpanzees, and orangutans) differed in their overall viewing patterns (the strength of...
viewing the objects vs. the agent), their pattern of predictive looks did not differ between the species. Thus, we conclude that online goal-based action prediction is not uniquely human, but are shared more widely among hominoids. The only potential difference between this and previous studies with infants is that, when viewing the claw actions during the test event, infants made location-based action predictions (i.e. more frequent first look to the distractor than to the target), while apes made no significant prediction (i.e. chance level). However, although not significant but consistent with a location-based prediction, apes tended to view the distractor longer than target object in the claw condition. Nevertheless, both location-based and chance-level prediction are theoretically plausible given that human infants did not distinguish the outcomes of the claw action in the looking-time paradigm (Woodward, 1998). Future studies are needed to confirm whether this potential difference reflects merely methodological issues or the species difference. In any case, the most important result of this and previous studies is that apes and infants made the goal-based prediction only in the hand condition, although the claw action similarly captured their attention when the claw/hand was grasping the target.

Why then do great apes seem to predict the goal of a hand but not a claw action? One possibility is that action familiarity shapes action understanding in great apes, as shown in many previous studies with macaque monkeys and humans (Falck-Ytter, et al., 2006; Flanagan & Johansson, 2003; Rochat, et al., 2008). For example, in human infants, the first-person motor experience (rather than the visual experience) of an unfamiliar tool-use action helped infants to understand the goal of that action (Sommerville, Hildebrand, & Crane, 2008). The possible mechanism underlying these phenomena is a direct matching process, in which observed actions are mapped onto motor representations of that action (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fogassi, & Gallese, 2001). Thus,
this same mechanism may be shared widely in phylogeny (Bonini & Ferrari, 2011; Hecht et al., 2013), facilitating the processing of familiar over unfamiliar actions in these species. However, the previous studies also suggest other possibilities. For example, providing infants with additional abstract movement cues helped them to understand the goal of mechanical object (Biro & Leslie, 2007; Luo & Baillargeon, 2005). Also, after seeing or interacting with a human agent operating the mechanical claw, infants understood the goal of mechanical claw (Gerson & Woodward, 2012; Hofer, Hauf, & Aschersleben, 2005). Future studies should consider these possibilities to further examine the underlying mechanisms for action understanding in great apes and the other nonhuman primates.

Does the current finding, the goal-based action prediction in great apes, reflect their understandings of mental or intentional state of others? Recent evidence and theories suggest that the answer is not necessarily yes (e.g. teleological stance theory; Gergely & Csibra, 2003). That is, they suggest that humans may develop two modes of action interpretations ontogenetically, in which the attribution of goal states to others may precede the attribution of mental states to others. These two modes of action interpretations can be distinguished from one another by examining whether the interpreted actions are based on the true belief or false belief. Thus, it is conceivable that human 11-month-old infants who showed the goal attribution in the current test (Cannon & Woodward, 2012) might fail to pass the nonverbal false-belief task (Southgate, et al., 2007). As the previous studies with great apes have consistently failed to produce positive evidence for nonverbal false-belief attribution (Call & Tomasello, 2008; Kaminski, Call, & Tomasello, 2008), future studies should examine whether they also fail to make online prediction about the other’s actions based on false beliefs.

In conclusion, by adopting an eye-tracking technique with great apes and the same
method from the previous study with infants, we found evidence that great apes also make online goal-based predictions about the other’s actions. In line with the other comparative studies, we suggest that humans are not the only hominoids who are sensitive to the other’s goal and spontaneously predict the other’s action.

Acknowledgement

This study was in part funded by the Japan Society for Promotion of Science for study abroad. We thank the keepers of the Wolfgang Köhler Primate Research Center for the help of data collection.

References


Comparative Psychology, 126(4), 388-398.


